

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 December 2002 (12.12.2002)

PCT

(10) International Publication Number
WO 02/098890 A2

(51) International Patent Classification⁷: C07H (US). LIOUBIN, Mario, N. [US/US]; 3014 Los Prados # A310, San Mateo, CA 94403 (US).

(21) International Application Number: PCT/US02/17461

(74) Agents: BRUNELLE, Jan et al.; Exelixis, Inc., P.O. Box 511, 170 Harbor Way, South San Francisco, CA 94083-0511 (US).

(22) International Filing Date: 3 June 2002 (03.06.2002)

(25) Filing Language: English

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(26) Publication Language: English

(30) Priority Data:
60/296,076 5 June 2001 (05.06.2001) US
60/328,605 10 October 2001 (10.10.2001) US
60/357,253 15 February 2002 (15.02.2002) US

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): EX-ELIXIS, INC. [US/US]; 170 Harbor Way, P.O. Box 511, South San Francisco, CA 94083-0511 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): FRIEDMAN, Lori [US/US]; One Bayside Village Place, Unit 212, San Francisco, CA 94107 (US). PLOWMAN, Gregory, D. [US/US]; 35 Winding Way, San Carlos, CA 94070 (US). BELVIN, Marcia [US/US]; 921 Santa Fe Avenue, Albany, CA 94706 (US). FRANCIS-LANG, Helen [GB/US]; 1782 Pacific Avenue, Apt. 2, San Francisco, CA 94109 (US). LI, Danxi [CN/US]; 90 Behr Avenue #302, San Francisco, CA 94131 (US). FUNKE, Roel, P. [NL/US]; 343 California Avenue, South San Francisco, CA 94080

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/098890 A2

(54) Title: MARKs AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE

(57) Abstract: Human MARK genes are identified as modulators of the p53 pathway, and thus are therapeutic targets for disorders associated with defective p53 function. Methods for identifying modulators of p53, comprising screening for agents that modulate the activity of MARK are provided.

MARKS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE**REFERENCE TO RELATED APPLICATIONS**

This application claims priority to U.S. provisional patent applications 60/296,076
5 filed 6/5/2001, 60/328,605 filed 10/10/2001, and 60/357,253 filed 2/15/2002. The
contents of the prior applications are hereby incorporated in their entirety.

BACKGROUND OF THE INVENTION

The p53 gene is mutated in over 50 different types of human cancers, including
10 familial and spontaneous cancers, and is believed to be the most commonly mutated gene
in human cancer (Zambetti and Levine, FASEB (1993) 7:855-865; Hollstein, *et al.*,
Nucleic Acids Res. (1994) 22:3551-3555). Greater than 90% of mutations in the p53 gene
are missense mutations that alter a single amino acid that inactivates p53 function.
Aberrant forms of human p53 are associated with poor prognosis, more aggressive tumors,
15 metastasis, and short survival rates (Mitsudomi *et al.*, Clin Cancer Res 2000 Oct;
6(10):4055-63; Koshland, Science (1993) 262:1953).

The human p53 protein normally functions as a central integrator of signals including
DNA damage, hypoxia, nucleotide deprivation, and oncogene activation (Prives, Cell
(1998) 95:5-8). In response to these signals, p53 protein levels are greatly increased with
20 the result that the accumulated p53 activates cell cycle arrest or apoptosis depending on
the nature and strength of these signals. Indeed, multiple lines of experimental evidence
have pointed to a key role for p53 as a tumor suppressor (Levine, Cell (1997) 88:323-331).
For example, homozygous p53 "knockout" mice are developmentally normal but exhibit
nearly 100% incidence of neoplasia in the first year of life (Donehower *et al.*, Nature
25 (1992) 356:215-221).

The biochemical mechanisms and pathways through which p53 functions in normal
and cancerous cells are not fully understood, but one clearly important aspect of p53
function is its activity as a gene-specific transcriptional activator. Among the genes with
known p53-response elements are several with well-characterized roles in either regulation
30 of the cell cycle or apoptosis, including GADD45, p21/Waf1/Cip1, cyclin G, Bax, IGF-
BP3, and MDM2 (Levine, Cell (1997) 88:323-331).

Microtubules have a central role in the regulation of cell shape and polarity during
differentiation, chromosome partitioning at mitosis, and intracellular transport.

Microtubules undergo rearrangements involving rapid transitions between stable and

dynamic states during these processes. Microtubule affinity regulating kinases (MARK) are a novel family of protein kinases that phosphorylate microtubule-associated proteins and trigger microtubule disruption (Drewes, G., et al. (1997) *Cell* 89: 297-308).

5 Microtubule affinity regulating kinase 1 (MARK1) is a serine/threonine kinase that phosphorylates microtubule-associated protein tau, leading to disruption of microtubules. It shares 90% amino acid homology with the rat version of MARK1, and demonstrates ubiquitous expression with highest levels in testis and brain (Nagase, T. et al. (2000) *DNA Res.* 7: 143-150).

EMK1 (MARK2) is a serine/threonine protein kinase with two isoforms, which differ 10 by the presence or absence of a 162-bp alternative exon (Espinosa, L. and Navarro, E. (1998) *Cytogenet. Cell Genet.* 81:278-282). Both human isoforms are coexpressed in a number of cell lines and tissues, with the highest expression found in heart, brain, placenta, skeletal muscle, and pancreas, and at lower levels in lung, liver, and kidney 15 (Inglis, J. et al. (1993) *Mammalian Genome* 4: 401-403). Due to the physical location of this gene, 11q12-q13, EMK1 is a candidate gene for carcinogenic events (Courseaux, A. et al. (1995) *Mammalian Genome* 6: 311-312), and has been associated with colon and prostate cancer (Moore, T. M., et al. (2000) *J Biol Chem* 275:4311-22; Navarro, E., et al. (1999) *Biochim Biophys Acta* 1450: 254-64).

20 Microtubule affinity regulating kinase 3 (MARK3) was originally identified as a marker (KP78) induced by treatment with DNA damaging agents. The loss of MARK3 was associated with carcinogenesis in the pancreas (Parsa, I. (1988) *Cell Growth Differ.* 9: 197-208). MARK3 may be involved in cell cycle regulation, and alterations in the MARK3 gene may lead to carcinogenesis. MARK 3 is ubiquitously expressed throughout 25 human tissues, with an additional 3.0 Kb transcript present in the heart (Peng, C. et al. (1998) *Cell Growth Differ.* 9: 197-208).

MAP/microtubule affinity-regulating kinase like 1 (MARKL1) has two isoforms (Nagase, T. et al. (2001) *DNA Res.* 8: 85-95), is activated by the beta-catenin/Tcf complex in hepatic cell lines, and may be involved in hepatic carcinogenesis (Kato, T. et al. (2001). *Neoplasia* 3:4-9).

30 The ability to manipulate the genomes of model organisms such as *Drosophila* provides a powerful means to analyze biochemical processes that, due to significant evolutionary conservation, has direct relevance to more complex vertebrate organisms. Due to a high level of gene and pathway conservation, the strong similarity of cellular processes, and the functional conservation of genes between these model organisms and

mammals, identification of the involvement of novel genes in particular pathways and their functions in such model organisms can directly contribute to the understanding of the correlative pathways and methods of modulating them in mammals (see, for example, Mechler BM et al., 1985 EMBO J 4:1551-1557; Gateff E. 1982 Adv. Cancer Res. 37: 33-5 74; Watson KL., et al., 1994 J Cell Sci. 18: 19-33; Miklos GL, and Rubin GM. 1996 Cell 86:521-529; Wassarman DA, et al., 1995 Curr Opin Gen Dev 5: 44-50; and Booth DR. 1999 Cancer Metastasis Rev. 18: 261-284). For example, a genetic screen can be carried out in an invertebrate model organism having underexpression (e.g. knockout) or overexpression of a gene (referred to as a "genetic entry point") that yields a visible 10 phenotype. Additional genes are mutated in a random or targeted manner. When a gene mutation changes the original phenotype caused by the mutation in the genetic entry point, the gene is identified as a "modifier" involved in the same or overlapping pathway as the genetic entry point. When the genetic entry point is an ortholog of a human gene implicated in a disease pathway, such as p53, modifier genes can be identified that may be 15 attractive candidate targets for novel therapeutics.

All references cited herein, including sequence information in referenced Genbank identifier numbers and website references, are incorporated herein in their entireties.

SUMMARY OF THE INVENTION

We have discovered genes that modify the p53 pathway in *Drosophila*, and identified 20 their human orthologs, hereinafter referred to as MARK. The invention provides methods for utilizing these p53 modifier genes and polypeptides to identify candidate therapeutic agents that can be used in the treatment of disorders associated with defective p53 function. Preferred MARK-modulating agents specifically bind to MARK polypeptides 25 and restore p53 function. Other preferred MARK-modulating agents are nucleic acid modulators such as antisense oligomers and RNAi that repress MARK gene expression or product activity by, for example, binding to and inhibiting the respective nucleic acid (i.e. DNA or mRNA).

MARK-specific modulating agents may be evaluated by any convenient *in vitro* or *in* 30 *vivo* assay for molecular interaction with a MARK polypeptide or nucleic acid. In one embodiment, candidate p53 modulating agents are tested with an assay system comprising a MARK polypeptide or nucleic acid. Candidate agents that produce a change in the activity of the assay system relative to controls are identified as candidate p53 modulating agents. The assay system may be cell-based or cell-free. MARK-modulating agents

include MARK related proteins (e.g. dominant negative mutants, and biotherapeutics); MARK-specific antibodies; MARK-specific antisense oligomers and other nucleic acid modulators; and chemical agents that specifically bind MARK or compete with MARK binding target. In one specific embodiment, a small molecule modulator is identified 5 using a kinase assay. In specific embodiments, the screening assay system is selected from a binding assay, an apoptosis assay, a cell proliferation assay, an angiogenesis assay, and a hypoxic induction assay.

In another embodiment, candidate p53 pathway modulating agents are further tested using a second assay system that detects changes in the p53 pathway, such as angiogenic, 10 apoptotic, or cell proliferation changes produced by the originally identified candidate agent or an agent derived from the original agent. The second assay system may use cultured cells or non-human animals. In specific embodiments, the secondary assay system uses non-human animals, including animals predetermined to have a disease or disorder implicating the p53 pathway, such as an angiogenic, apoptotic, or cell 15 proliferation disorder (e.g. cancer).

The invention further provides methods for modulating the p53 pathway in a mammalian cell by contacting the mammalian cell with an agent that specifically binds a MARK polypeptide or nucleic acid. The agent may be a small molecule modulator, a nucleic acid modulator, or an antibody and may be administered to a mammalian animal 20 predetermined to have a pathology associated the p53 pathway.

DETAILED DESCRIPTION OF THE INVENTION

Genetic screens were designed to identify modifiers of the p53 pathway in *Drosophila* in which p53 was overexpressed in the wing (Ollmann M, et al., Cell 2000 101: 91-101). 25 The KP78a gene was identified as a modifier of the p53 pathway. Accordingly, vertebrate orthologs of these modifiers, and preferably the human orthologs, microtubule affinity regulator kinase (MARK) genes (i.e., nucleic acids and polypeptides) are attractive drug targets for the treatment of pathologies associated with a defective p53 signaling pathway, such as cancer.

30 In vitro and in vivo methods of assessing MARK function are provided herein. Modulation of the MARK or their respective binding partners is useful for understanding the association of the p53 pathway and its members in normal and disease conditions and for developing diagnostics and therapeutic modalities for p53 related pathologies. MARK-modulating agents that act by inhibiting or enhancing MARK expression, directly

or indirectly, for example, by affecting a MARK function such as enzymatic (e.g., catalytic) or binding activity, can be identified using methods provided herein. MARK modulating agents are useful in diagnosis, therapy and pharmaceutical development.

5 **Nucleic acids and polypeptides of the invention**

Sequences related to MARK nucleic acids and polypeptides that can be used in the invention are disclosed in Genbank (referenced by Genbank identifier (GI) number) as GI#s 9845486 (SEQ ID NO:1), 9845488 (SEQ ID NO:2), 18578044 (SEQ ID NO:3), 14250621 (SEQ ID NO:6), 15042610 (SEQ ID NO:7), 8923921 (SEQ ID NO:8), 10 17445805 (SEQ ID NO:9), 7959214 (SEQ ID NO:11), 14042208 (SEQ ID NO:12), 3089348 (SEQ ID NO:13), 4505102 (SEQ ID NO:14), 5714635 (SEQ ID NO:15), 15 18448970 (SEQ ID NO:18), 13366083 (SEQ ID NO:19), 14017936 (SEQ ID NO:22), and 16555377 (SEQ ID NO:23) for nucleic acid, and GI#s 9845487 (SEQ ID NO:24), 8923922 (SEQ ID NO:25), 3089349 (SEQ ID NO:26), 4505103 (SEQ ID NO:27), 13366084 (SEQ ID NO:28) and 13899225 (SEQ ID NO:29) for polypeptides.

Additionally, nucleic acid sequences of SEQ ID NOs:4, 5, 16, 17, 20, 21, and novel nucleic acid sequence of SEQ ID NO:10 can also be used in the invention.

MARKs are kinase proteins with kinase and UBA/TS-N domains. The term "MARK polypeptide" refers to a full-length MARK protein or a functionally active fragment or derivative thereof. A "functionally active" MARK fragment or derivative exhibits one or more functional activities associated with a full-length, wild-type MARK protein, such as antigenic or immunogenic activity, enzymatic activity, ability to bind natural cellular substrates, etc. The functional activity of MARK proteins, derivatives and fragments can be assayed by various methods known to one skilled in the art (Current Protocols in Protein Science (1998) Coligan *et al.*, eds., John Wiley & Sons, Inc., Somerset, New Jersey) and as further discussed below. For purposes herein, functionally active fragments also include those fragments that comprise one or more structural domains of a MARK, such as a kinase domain or a binding domain. Protein domains can be identified using the PFAM program (Bateman A., et al., Nucleic Acids Res, 1999, 27:260-2; 30 <http://pfam.wustl.edu>). For example, the protein kinase domains of MARKs from GI#s 9845487 (SEQ ID NO:24), 8923922 (SEQ ID NO:25), 4505103 (SEQ ID NO:27), and 13899225 (SEQ ID NO:29) is located at approximately amino acid residues 20 to 271, 60 to 311, 56 to 307, and 59 to 310, respectively (PFAM 00069). Further, the ubiquitin associated (UBA/TS-N) domains of MARKs from GI#s 9845487 (SEQ ID NO:24),

8923922 (SEQ ID NO:25), 4505103 (SEQ ID NO:27 and 13899225 (SEQ ID NO:29) is located at approximately amino acid residues 291 to 330, 331 to 370, 327 to 366, and 330 to 369, respectively (PFAM 00627). Methods for obtaining MARK polypeptides are also further described below. In some embodiments, preferred fragments are functionally active, domain-containing fragments comprising at least 25 contiguous amino acids, preferably at least 50, more preferably 75, and most preferably at least 100 contiguous amino acids of any one of SEQ ID NOs:24, 25, 26, 27, 28, or 29 (a MARK). In further preferred embodiments, the fragment comprises the entire kinase (functionally active) domain.

The term "MARK nucleic acid" refers to a DNA or RNA molecule that encodes a MARK polypeptide. Preferably, the MARK polypeptide or nucleic acid or fragment thereof is from a human, but can also be an ortholog, or derivative thereof with at least 70% sequence identity, preferably at least 80%, more preferably 85%, still more preferably 90%, and most preferably at least 95% sequence identity with MARK.

Normally, orthologs in different species retain the same function, due to presence of one or more protein motifs and/or 3-dimensional structures. Orthologs are generally identified by sequence homology analysis, such as BLAST analysis, usually using protein bait sequences. Sequences are assigned as a potential ortholog if the best hit sequence from the forward BLAST result retrieves the original query sequence in the reverse BLAST (Huynen MA and Bork P, Proc Natl Acad Sci (1998) 95:5849-5856; Huynen MA *et al.*, Genome Research (2000) 10:1204-1210). Programs for multiple sequence alignment, such as CLUSTAL (Thompson JD *et al.*, 1994, Nucleic Acids Res 22:4673-4680) may be used to highlight conserved regions and/or residues of orthologous proteins and to generate phylogenetic trees. In a phylogenetic tree representing multiple homologous sequences from diverse species (e.g., retrieved through BLAST analysis), orthologous sequences from two species generally appear closest on the tree with respect to all other sequences from these two species. Structural threading or other analysis of protein folding (e.g., using software by ProCeryon, Biosciences, Salzburg, Austria) may also identify potential orthologs. In evolution, when a gene duplication event follows speciation, a single gene in one species, such as *Drosophila*, may correspond to multiple genes (paralogs) in another, such as human. As used herein, the term "orthologs" encompasses paralogs. As used herein, "percent (%) sequence identity" with respect to a subject sequence, or a specified portion of a subject sequence, is defined as the percentage of nucleotides or amino acids in the candidate derivative sequence identical with the nucleotides or amino acids in the

subject sequence (or specified portion thereof), after aligning the sequences and introducing gaps, if necessary to achieve the maximum percent sequence identity, as generated by the program WU-BLAST-2.0a19 (Altschul *et al.*, J. Mol. Biol. (1997) 215:403-410; <http://blast.wustl.edu/blast/README.html>) with all the search parameters set to default values. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched. A % identity value is determined by the number of matching identical nucleotides or amino acids divided by the sequence length for which the percent identity is being reported. "Percent (%) amino acid sequence similarity" is determined by doing the same calculation as for determining % amino acid sequence identity, but including conservative amino acid substitutions in addition to identical amino acids in the computation.

A conservative amino acid substitution is one in which an amino acid is substituted for another amino acid having similar properties such that the folding or activity of the protein is not significantly affected. Aromatic amino acids that can be substituted for each other are phenylalanine, tryptophan, and tyrosine; interchangeable hydrophobic amino acids are leucine, isoleucine, methionine, and valine; interchangeable polar amino acids are glutamine and asparagine; interchangeable basic amino acids are arginine, lysine and histidine; interchangeable acidic amino acids are aspartic acid and glutamic acid; and interchangeable small amino acids are alanine, serine, threonine, cysteine and glycine.

Alternatively, an alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman (Smith and Waterman, 1981, Advances in Applied Mathematics 2:482-489; database: European Bioinformatics Institute <http://www.ebi.ac.uk/MPsrch/>; Smith and Waterman, 1981, J. of Molec.Biol., 147:195-197; Nicholas *et al.*, 1998, "A Tutorial on Searching Sequence Databases and Sequence Scoring Methods" (www.psc.edu) and references cited therein.; W.R. Pearson, 1991, Genomics 11:635-650). This algorithm can be applied to amino acid sequences by using the scoring matrix developed by Dayhoff (Dayhoff: Atlas of Protein Sequences and Structure, M. O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA), and normalized by Gribskov (Gribskov 1986 Nucl. Acids Res. 14(6):6745-6763). The Smith-Waterman algorithm may be employed where default parameters are used for scoring (for example, gap open penalty of 12, gap

extension penalty of two). From the data generated, the "Match" value reflects "sequence identity."

- Derivative nucleic acid molecules of the subject nucleic acid molecules include sequences that hybridize to the nucleic acid sequence of any of SEQ ID NOs:1 through 23.
- 5 The stringency of hybridization can be controlled by temperature, ionic strength, pH, and the presence of denaturing agents such as formamide during hybridization and washing. Conditions routinely used are set out in readily available procedure texts (e.g., Current Protocol in Molecular Biology, Vol. 1, Chap. 2.10, John Wiley & Sons, Publishers (1994); Sambrook *et al.*, Molecular Cloning, Cold Spring Harbor (1989)). In some embodiments,
- 10 a nucleic acid molecule of the invention is capable of hybridizing to a nucleic acid molecule containing the nucleotide sequence of any one of SEQ ID NOs:1 through 23 under stringent hybridization conditions that comprise: prehybridization of filters containing nucleic acid for 8 hours to overnight at 65° C in a solution comprising 6X single strength citrate (SSC) (1X SSC is 0.15 M NaCl, 0.015 M Na citrate; pH 7.0), 5X
- 15 Denhardt's solution, 0.05% sodium pyrophosphate and 100 µg/ml herring sperm DNA; hybridization for 18-20 hours at 65° C in a solution containing 6X SSC, 1X Denhardt's solution, 100 µg/ml yeast tRNA and 0.05% sodium pyrophosphate; and washing of filters at 65° C for 1h in a solution containing 0.2X SSC and 0.1% SDS (sodium dodecyl sulfate).
- 20 In other embodiments, moderately stringent hybridization conditions are used that comprise: pretreatment of filters containing nucleic acid for 6 h at 40° C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH7.5), 5mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 µg/ml denatured salmon sperm DNA; hybridization for 18-20h at 40° C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH7.5), 5mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 µg/ml salmon sperm DNA, and 10% (wt/vol) dextran sulfate; followed by washing twice for 1 hour at 55° C in a solution containing 2X SSC and 0.1% SDS.
- 25 Alternatively, low stringency conditions can be used that comprise: incubation for 8 hours to overnight at 37° C in a solution comprising 20% formamide, 5 x SSC, 50 mM sodium phosphate (pH 7.6), 5X Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured sheared salmon sperm DNA; hybridization in the same buffer for 18 to 20 hours; and washing of filters in 1 x SSC at about 37° C for 1 hour.

Isolation, Production, Expression, and Mis-expression of MARK Nucleic Acids and Polypeptides

MARK nucleic acids and polypeptides, useful for identifying and testing agents that modulate MARK function and for other applications related to the involvement of MARK in the p53 pathway. MARK nucleic acids and derivatives and orthologs thereof may be obtained using any available method. For instance, techniques for isolating cDNA or genomic DNA sequences of interest by screening DNA libraries or by using polymerase chain reaction (PCR) are well known in the art. In general, the particular use for the protein will dictate the particulars of expression, production, and purification methods.

For instance, production of proteins for use in screening for modulating agents may require methods that preserve specific biological activities of these proteins, whereas production of proteins for antibody generation may require structural integrity of particular epitopes. Expression of proteins to be purified for screening or antibody production may require the addition of specific tags (*e.g.*, generation of fusion proteins). Overexpression of a MARK protein for assays used to assess MARK function, such as involvement in cell cycle regulation or hypoxic response, may require expression in eukaryotic cell lines capable of these cellular activities. Techniques for the expression, production, and purification of proteins are well known in the art; any suitable means therefore may be used (*e.g.*, Higgins SJ and Hames BD (eds.) *Protein Expression: A Practical Approach*, Oxford University Press Inc., New York 1999; Stanbury PF et al., *Principles of Fermentation Technology*, 2nd edition, Elsevier Science, New York, 1995; Doonan S (ed.) *Protein Purification Protocols*, Humana Press, New Jersey, 1996; Coligan JE et al, *Current Protocols in Protein Science* (eds.), 1999, John Wiley & Sons, New York). In particular embodiments, recombinant MARK is expressed in a cell line known to have defective p53 function (*e.g.* SAOS-2 osteoblasts, H1299 lung cancer cells, C33A and HT3 cervical cancer cells, HT-29 and DLD-1 colon cancer cells, among others, available from American Type Culture Collection (ATCC), Manassas, VA). The recombinant cells are used in cell-based screening assay systems of the invention, as described further below.

The nucleotide sequence encoding a MARK polypeptide can be inserted into any appropriate expression vector. The necessary transcriptional and translational signals, including promoter/enhancer element, can derive from the native MARK gene and/or its flanking regions or can be heterologous. A variety of host-vector expression systems may be utilized, such as mammalian cell systems infected with virus (*e.g.* vaccinia virus, adenovirus, *etc.*); insect cell systems infected with virus (*e.g.* baculovirus);

microorganisms such as yeast containing yeast vectors, or bacteria transformed with bacteriophage, plasmid, or cosmid DNA. A host cell strain that modulates the expression of, modifies, and/or specifically processes the gene product may be used.

To detect expression of the MARK gene product, the expression vector can comprise a promoter operably linked to a MARK gene nucleic acid, one or more origins of replication, and, one or more selectable markers (*e.g.* thymidine kinase activity, resistance to antibiotics, *etc.*). Alternatively, recombinant expression vectors can be identified by assaying for the expression of the MARK gene product based on the physical or functional properties of the MARK protein in *in vitro* assay systems (*e.g.* immunoassays).

The MARK protein, fragment, or derivative may be optionally expressed as a fusion, or chimeric protein product (*i.e.* it is joined via a peptide bond to a heterologous protein sequence of a different protein), for example to facilitate purification or detection. A chimeric product can be made by ligating the appropriate nucleic acid sequences encoding the desired amino acid sequences to each other using standard methods and expressing the chimeric product. A chimeric product may also be made by protein synthetic techniques, *e.g.* by use of a peptide synthesizer (Hunkapiller *et al.*, *Nature* (1984) 310:105-111).

Once a recombinant cell that expresses the MARK gene sequence is identified, the gene product can be isolated and purified using standard methods (*e.g.* ion exchange, affinity, and gel exclusion chromatography; centrifugation; differential solubility; electrophoresis, cite purification reference). Alternatively, native MARK proteins can be purified from natural sources, by standard methods (*e.g.* immunoaffinity purification). Once a protein is obtained, it may be quantified and its activity measured by appropriate methods, such as immunoassay, bioassay, or other measurements of physical properties, such as crystallography.

The methods of this invention may also use cells that have been engineered for altered expression (mis-expression) of MARK or other genes associated with the p53 pathway. As used herein, mis-expression encompasses ectopic expression, over-expression, under-expression, and non-expression (*e.g.* by gene knock-out or blocking expression that would otherwise normally occur).

30

Genetically modified animals

Animal models that have been genetically modified to alter MARK expression may be used in *in vivo* assays to test for activity of a candidate p53 modulating agent, or to further assess the role of MARK in a p53 pathway process such as apoptosis or cell proliferation.

Preferably, the altered MARK expression results in a detectable phenotype, such as decreased or increased levels of cell proliferation, angiogenesis, or apoptosis compared to control animals having normal MARK expression. The genetically modified animal may additionally have altered p53 expression (e.g. p53 knockout). Preferred genetically modified animals are mammals such as primates, rodents (preferably mice), cows, horses, goats, sheep, pigs, dogs and cats. Preferred non-mammalian species include zebrafish, *C. elegans*, and *Drosophila*. Preferred genetically modified animals are transgenic animals having a heterologous nucleic acid sequence present as an extrachromosomal element in a portion of its cells, i.e. mosaic animals (see, for example, techniques described by Jakobovits, 1994, *Curr. Biol.* 4:761-763.) or stably integrated into its germ line DNA (i.e., in the genomic sequence of most or all of its cells). Heterologous nucleic acid is introduced into the germ line of such transgenic animals by genetic manipulation of, for example, embryos or embryonic stem cells of the host animal.

Methods of making transgenic animals are well-known in the art (for transgenic mice see Brinster et al., *Proc. Nat. Acad. Sci. USA* 82: 4438-4442 (1985), U.S. Pat. Nos. 4,736,866 and 4,870,009, both by Leder et al., U.S. Pat. No. 4,873,191 by Wagner et al., and Hogan, B., *Manipulating the Mouse Embryo*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., (1986); for particle bombardment see U.S. Pat. No., 4,945,050, by Sandford et al.; for transgenic *Drosophila* see Rubin and Spradling, *Science* (1982) 218:348-53 and U.S. Pat. No. 4,670,388; for transgenic insects see Berghammer A.J. et al., *A Universal Marker for Transgenic Insects* (1999) *Nature* 402:370-371; for transgenic Zebrafish see Lin S., *Transgenic Zebrafish*, *Methods Mol Biol.* (2000);136:375-3830); for microinjection procedures for fish, amphibian eggs and birds see Houdebine and Chourrout, *Experientia* (1991) 47:897-905; for transgenic rats see Hammer et al., *Cell* (1990) 63:1099-1112; and for culturing of embryonic stem (ES) cells and the subsequent production of transgenic animals by the introduction of DNA into ES cells using methods such as electroporation, calcium phosphate/DNA precipitation and direct injection see, e.g., *Teratocarcinomas and Embryonic Stem Cells, A Practical Approach*, E. J. Robertson, ed., IRL Press (1987)). Clones of the nonhuman transgenic animals can be produced according to available methods (see Wilmut, I. et al. (1997) *Nature* 385:810-813; and PCT International Publication Nos. WO 97/07668 and WO 97/07669).

In one embodiment, the transgenic animal is a "knock-out" animal having a heterozygous or homozygous alteration in the sequence of an endogenous MARK gene that results in a decrease of MARK function, preferably such that MARK expression is

undetectable or insignificant. Knock-out animals are typically generated by homologous recombination with a vector comprising a transgene having at least a portion of the gene to be knocked out. Typically a deletion, addition or substitution has been introduced into the transgene to functionally disrupt it. The transgene can be a human gene (e.g., from a 5 human genomic clone) but more preferably is an ortholog of the human gene derived from the transgenic host species. For example, a mouse MARK gene is used to construct a homologous recombination vector suitable for altering an endogenous MARK gene in the mouse genome. Detailed methodologies for homologous recombination in mice are available (see Capecchi, *Science* (1989) 244:1288-1292; Joyner *et al.*, *Nature* (1989) 10 338:153-156). Procedures for the production of non-rodent transgenic mammals and other animals are also available (Houdebine and Chourrout, *supra*; Pursel *et al.*, *Science* (1989) 244:1281-1288; Simms *et al.*, *Bio/Technology* (1988) 6:179-183). In a preferred embodiment, knock-out animals, such as mice harboring a knockout of a specific gene, may be used to produce antibodies against the human counterpart of the gene that has been 15 knocked out (Claesson MH *et al.*, (1994) *Scan J Immunol* 40:257-264; Declerck PJ *et al.*, (1995) *J Biol Chem*. 270:8397-400).

In another embodiment, the transgenic animal is a "knock-in" animal having an alteration in its genome that results in altered expression (e.g., increased (including ectopic) or decreased expression) of the MARK gene, e.g., by introduction of additional 20 copies of MARK, or by operatively inserting a regulatory sequence that provides for altered expression of an endogenous copy of the MARK gene. Such regulatory sequences include inducible, tissue-specific, and constitutive promoters and enhancer elements. The knock-in can be homozygous or heterozygous.

Transgenic nonhuman animals can also be produced that contain selected systems 25 allowing for regulated expression of the transgene. One example of such a system that may be produced is the cre/loxP recombinase system of bacteriophage P1 (Lakso *et al.*, *PNAS* (1992) 89:6232-6236; U.S. Pat. No. 4,959,317). If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be 30 provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355; U.S. Pat. No. 5,654,182). In a preferred embodiment,

both Cre-LoxP and Flp-Frt are used in the same system to regulate expression of the transgene, and for sequential deletion of vector sequences in the same cell (Sun X et al (2000) *Nat Genet* 25:83-6).

The genetically modified animals can be used in genetic studies to further elucidate the p53 pathway, as animal models of disease and disorders implicating defective p53 function, and for *in vivo* testing of candidate therapeutic agents, such as those identified in screens described below. The candidate therapeutic agents are administered to a genetically modified animal having altered MARK function and phenotypic changes are compared with appropriate control animals such as genetically modified animals that receive placebo treatment, and/or animals with unaltered MARK expression that receive candidate therapeutic agent.

In addition to the above-described genetically modified animals having altered MARK function, animal models having defective p53 function (and otherwise normal MARK function), can be used in the methods of the present invention. For example, a p53 knockout mouse can be used to assess, *in vivo*, the activity of a candidate p53 modulating agent identified in one of the *in vitro* assays described below. p53 knockout mice are described in the literature (Jacks et al., *Nature* 2001;410:1111-1116, 1043-1044; Donehower et al., supra). Preferably, the candidate p53 modulating agent when administered to a model system with cells defective in p53 function, produces a detectable phenotypic change in the model system indicating that the p53 function is restored, i.e., the cells exhibit normal cell cycle progression.

Modulating Agents

The invention provides methods to identify agents that interact with and/or modulate the function of MARK and/or the p53 pathway. Such agents are useful in a variety of diagnostic and therapeutic applications associated with the p53 pathway, as well as in further analysis of the MARK protein and its contribution to the p53 pathway. Accordingly, the invention also provides methods for modulating the p53 pathway comprising the step of specifically modulating MARK activity by administering a MARK-interacting or -modulating agent.

In a preferred embodiment, MARK-modulating agents inhibit or enhance MARK activity or otherwise affect normal MARK function, including transcription, protein expression, protein localization, and cellular or extra-cellular activity. In a further preferred embodiment, the candidate p53 pathway- modulating agent specifically

modulates the function of the MARK. The phrases "specific modulating agent", "specifically modulates", etc., are used herein to refer to modulating agents that directly bind to the MARK polypeptide or nucleic acid, and preferably inhibit, enhance, or otherwise alter, the function of the MARK. The term also encompasses modulating agents that alter the interaction of the MARK with a binding partner or substrate (e.g. by binding to a binding partner of a MARK, or to a protein/binding partner complex, and inhibiting function).

Preferred MARK-modulating agents include small molecule compounds; MARK-interacting proteins, including antibodies and other biotherapeutics; and nucleic acid modulators such as antisense and RNA inhibitors. The modulating agents may be formulated in pharmaceutical compositions, for example, as compositions that may comprise other active ingredients, as in combination therapy, and/or suitable carriers or excipients. Techniques for formulation and administration of the compounds may be found in "Remington's Pharmaceutical Sciences" Mack Publishing Co., Easton, PA, 19th edition.

Small molecule modulators

Small molecules, are often preferred to modulate function of proteins with enzymatic function, and/or containing protein interaction domains. Chemical agents, referred to in the art as "small molecule" compounds are typically organic, non-peptide molecules, having a molecular weight less than 10,000, preferably less than 5,000, more preferably less than 1,000, and most preferably less than 500. This class of modulators includes chemically synthesized molecules, for instance, compounds from combinatorial chemical libraries. Synthetic compounds may be rationally designed or identified based on known or inferred properties of the MARK protein or may be identified by screening compound libraries. Alternative appropriate modulators of this class are natural products, particularly secondary metabolites from organisms such as plants or fungi, which can also be identified by screening compound libraries for MARK-modulating activity. Methods for generating and obtaining compounds are well known in the art (Schreiber SL, Science 2000) 151: 1964-1969; Radmann J and Gunther J, Science (2000) 151:1947-1948).

Small molecule modulators identified from screening assays, as described below, can be used as lead compounds from which candidate clinical compounds may be designed, optimized, and synthesized. Such clinical compounds may have utility in treating pathologies associated with the p53 pathway. The activity of candidate small molecule

modulating agents may be improved several-fold through iterative secondary functional validation, as further described below, structure determination, and candidate modulator modification and testing. Additionally, candidate clinical compounds are generated with specific regard to clinical and pharmacological properties. For example, the reagents may 5 be derivatized and re-screened using *in vitro* and *in vivo* assays to optimize activity and minimize toxicity for pharmaceutical development.

Protein Modulators

Specific MARK-interacting proteins are useful in a variety of diagnostic and 10 therapeutic applications related to the p53 pathway and related disorders, as well as in validation assays for other MARK-modulating agents. In a preferred embodiment, MARK-interacting proteins affect normal MARK function, including transcription, protein expression, protein localization, and cellular or extra-cellular activity. In another embodiment, MARK-interacting proteins are useful in detecting and providing 15 information about the function of MARK proteins, as is relevant to p53 related disorders, such as cancer (e.g., for diagnostic means).

An MARK-interacting protein may be endogenous, i.e. one that naturally interacts genetically or biochemically with a MARK, such as a member of the MARK pathway that modulates MARK expression, localization, and/or activity. MARK-modulators include 20 dominant negative forms of MARK-interacting proteins and of MARK proteins themselves. Yeast two-hybrid and variant screens offer preferred methods for identifying endogenous MARK-interacting proteins (Finley, R. L. et al. (1996) in DNA Cloning-Expression Systems: A Practical Approach, eds. Glover D. & Hames B. D (Oxford University Press, Oxford, England), pp. 169-203; Fashema SF et al., Gene (2000) 250:1- 25 14; Drees BL Curr Opin Chem Biol (1999) 3:64-70; Vidal M and Legrain P Nucleic Acids Res (1999) 27:919-29; and U.S. Pat. No. 5,928,868). Mass spectrometry is an alternative preferred method for the elucidation of protein complexes (reviewed in, e.g., Pandley A and Mann M, Nature (2000) 405:837-846; Yates JR 3rd, Trends Genet (2000) 16:5-8).

An MARK-interacting protein may be an exogenous protein, such as a MARK- 30 specific antibody or a T-cell antigen receptor (see, e.g., Harlow and Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor Laboratory; Harlow and Lane (1999) Using antibodies: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press). MARK antibodies are further discussed below.

In preferred embodiments, a MARK-interacting protein specifically binds a MARK protein. In alternative preferred embodiments, a MARK-modulating agent binds a MARK substrate, binding partner, or cofactor.

5 *Antibodies*

In another embodiment, the protein modulator is a MARK specific antibody agonist or antagonist. The antibodies have therapeutic and diagnostic utilities, and can be used in screening assays to identify MARK modulators. The antibodies can also be used in dissecting the portions of the MARK pathway responsible for various cellular responses 10 and in the general processing and maturation of the MARK.

Antibodies that specifically bind MARK polypeptides can be generated using known methods. Preferably the antibody is specific to a mammalian ortholog of MARK polypeptide, and more preferably, to human MARK. Antibodies may be polyclonal, monoclonal (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab 15 fragments, F(ab').sub.2 fragments, fragments produced by a FAb expression library, anti-idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above. Epitopes of MARK which are particularly antigenic can be selected, for example, by routine screening of MARK polypeptides for antigenicity or by applying a theoretical method for selecting antigenic regions of a protein (Hopp and Wood (1981), Proc. Nati. Acad. Sci. U.S.A. 78:3824-28; Hopp and Wood, (1983) Mol. Immunol. 20:483-89; 20 Sutcliffe et al., (1983) Science 219:660-66) to the amino acid sequence shown in any of SEQ ID NOs:24, 25, 26, 27, 28, or 29. Monoclonal antibodies with affinities of 10^8 M^{-1} preferably 10^9 M^{-1} to 10^{10} M^{-1} , or stronger can be made by standard procedures as described (Harlow and Lane, *supra*; Goding (1986) Monoclonal Antibodies: Principles 25 and Practice (2d ed) Academic Press, New York; and U.S. Pat. Nos. 4,381,292; 4,451,570; and 4,618,577). Antibodies may be generated against crude cell extracts of MARK or substantially purified fragments thereof. If MARK fragments are used, they preferably comprise at least 10, and more preferably, at least 20 contiguous amino acids of a MARK 30 protein. In a particular embodiment, MARK-specific antigens and/or immunogens are coupled to carrier proteins that stimulate the immune response. For example, the subject polypeptides are covalently coupled to the keyhole limpet hemocyanin (KLH) carrier, and the conjugate is emulsified in Freund's complete adjuvant, which enhances the immune response. An appropriate immune system such as a laboratory rabbit or mouse is immunized according to conventional protocols.

The presence of MARK-specific antibodies is assayed by an appropriate assay such as a solid phase enzyme-linked immunosorbant assay (ELISA) using immobilized corresponding MARK polypeptides. Other assays, such as radioimmunoassays or fluorescent assays might also be used.

5 Chimeric antibodies specific to MARK polypeptides can be made that contain different portions from different animal species. For instance, a human immunoglobulin constant region may be linked to a variable region of a murine mAb, such that the antibody derives its biological activity from the human antibody, and its binding specificity from the murine fragment. Chimeric antibodies are produced by splicing
10 together genes that encode the appropriate regions from each species (Morrison et al., Proc. Natl. Acad. Sci. (1984) 81:6851-6855; Neuberger et al., Nature (1984) 312:604-608; Takeda et al., Nature (1985) 31:452-454). Humanized antibodies, which are a form of chimeric antibodies, can be generated by grafting complementary-determining regions (CDRs) (Carlos, T. M., J. M. Harlan. 1994. Blood 84:2068-2101) of mouse antibodies
15 into a background of human framework regions and constant regions by recombinant DNA technology (Riechmann LM, et al., 1988 Nature 323: 323-327). Humanized antibodies contain ~10% murine sequences and ~90% human sequences, and thus further reduce or eliminate immunogenicity, while retaining the antibody specificities (Co MS, and Queen C. 1991 Nature 351: 501-501; Morrison SL. 1992 Ann. Rev. Immun.
20 10:239-265). Humanized antibodies and methods of their production are well-known in the art (U.S. Pat. Nos. 5,530,101, 5,585,089, 5,693,762, and 6,180,370).

MARK-specific single chain antibodies which are recombinant, single chain polypeptides formed by linking the heavy and light chain fragments of the Fv regions via an amino acid bridge, can be produced by methods known in the art (U.S. Pat. No. 25 4,946,778; Bird, Science (1988) 242:423-426; Huston et al., Proc. Natl. Acad. Sci. USA (1988) 85:5879-5883; and Ward et al., Nature (1989) 334:544-546).

Other suitable techniques for antibody production involve in vitro exposure of lymphocytes to the antigenic polypeptides or alternatively to selection of libraries of antibodies in phage or similar vectors (Huse et al., Science (1989) 246:1275-1281). As used herein, T-cell antigen receptors are included within the scope of antibody modulators (Harlow and Lane, 1988, *supra*).

The polypeptides and antibodies of the present invention may be used with or without modification. Frequently, antibodies will be labeled by joining, either covalently or non-covalently, a substance that provides for a detectable signal, or that is toxic to cells that

express the targeted protein (Menard S, et al., Int J. Biol Markers (1989) 4:131-134). A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent moieties, fluorescent emitting lanthanide metals, chemiluminescent moieties, bioluminescent moieties, magnetic particles, and the like (U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241). Also, recombinant immunoglobulins may be produced (U.S. Pat. No. 4,816,567). Antibodies to cytoplasmic polypeptides may be delivered and reach their targets by conjugation with membrane-penetrating toxin proteins (U.S. Pat. No. 10 6,086,900).

When used therapeutically in a patient, the antibodies of the subject invention are typically administered parenterally, when possible at the target site, or intravenously. The therapeutically effective dose and dosage regimen is determined by clinical studies. Typically, the amount of antibody administered is in the range of about 0.1 mg/kg –to 15 about 10 mg/kg of patient weight. For parenteral administration, the antibodies are formulated in a unit dosage injectable form (e.g., solution, suspension, emulsion) in association with a pharmaceutically acceptable vehicle. Such vehicles are inherently nontoxic and non-therapeutic. Examples are water, saline, Ringer's solution, dextrose solution, and 5% human serum albumin. Nonaqueous vehicles such as fixed oils, ethyl 20 oleate, or liposome carriers may also be used. The vehicle may contain minor amounts of additives, such as buffers and preservatives, which enhance isotonicity and chemical stability or otherwise enhance therapeutic potential. The antibodies' concentrations in such vehicles are typically in the range of about 1 mg/ml to about 10 mg/ml. Immunotherapeutic methods are further described in the literature (US Pat. No. 5,859,206; 25 WO0073469).

Nucleic Acid Modulators

Other preferred MARK-modulating agents comprise nucleic acid molecules, such as antisense oligomers or double stranded RNA (dsRNA), which generally inhibit MARK 30 activity. Preferred nucleic acid modulators interfere with the function of the MARK nucleic acid such as DNA replication, transcription, translocation of the MARK RNA to the site of protein translation, translation of protein from the MARK RNA, splicing of the MARK RNA to yield one or more mRNA species, or catalytic activity which may be engaged in or facilitated by the MARK RNA.

In one embodiment, the antisense oligomer is an oligonucleotide that is sufficiently complementary to a MARK mRNA to bind to and prevent translation, preferably by binding to the 5' untranslated region. MARK-specific antisense oligonucleotides, preferably range from at least 6 to about 200 nucleotides. In some embodiments the 5 oligonucleotide is preferably at least 10, 15, or 20 nucleotides in length. In other embodiments, the oligonucleotide is preferably less than 50, 40, or 30 nucleotides in length. The oligonucleotide can be DNA or RNA or a chimeric mixture or derivatives or modified versions thereof, single-stranded or double-stranded. The oligonucleotide can be modified at the base moiety, sugar moiety, or phosphate backbone. The oligonucleotide 10 may include other appending groups such as peptides, agents that facilitate transport across the cell membrane, hybridization-triggered cleavage agents, and intercalating agents.

In another embodiment, the antisense oligomer is a phosphothioate morpholino oligomer (PMO). PMOs are assembled from four different morpholino subunits, each of 15 which contain one of four genetic bases (A, C, G, or T) linked to a six-membered morpholine ring. Polymers of these subunits are joined by non-ionic phosphodiamidate intersubunit linkages. Details of how to make and use PMOs and other antisense oligomers are well known in the art (e.g. see WO99/18193; Probst JC, Antisense Oligodeoxynucleotide and Ribozyme Design, Methods. (2000) 22(3):271-281; Summerton 20 J, and Weller D. 1997 Antisense Nucleic Acid Drug Dev. :7:187-95; US Pat. No. 5,235,033; and US Pat No. 5,378,841).

Alternative preferred MARK nucleic acid modulators are double-stranded RNA species mediating RNA interference (RNAi). RNAi is the process of sequence-specific, post-transcriptional gene silencing in animals and plants, initiated by double-stranded 25 RNA (dsRNA) that is homologous in sequence to the silenced gene. Methods relating to the use of RNAi to silence genes in *C. elegans*, *Drosophila*, plants, and humans are known in the art (Fire A, et al., 1998 Nature 391:806-811; Fire, A. Trends Genet. 15, 358-363 (1999); Sharp, P. A. RNA interference 2001. Genes Dev. 15, 485-490 (2001); Hammond, S. M., et al., Nature Rev. Genet. 2, 110-1119 (2001); Tuschl, T. Chem. Biochem. 2, 239-30 245 (2001); Hamilton, A. et al., Science 286, 950-952 (1999); Hammond, S. M., et al., Nature 404, 293-296 (2000); Zamore, P. D., et al., Cell 101, 25-33 (2000); Bernstein, E., et al., Nature 409, 363-366 (2001); Elbashir, S. M., et al., Genes Dev. 15, 188-200 (2001); WO0129058; WO9932619; Elbashir SM, et al., 2001 Nature 411:494-498).

Nucleic acid modulators are commonly used as research reagents, diagnostics, and therapeutics. For example, antisense oligonucleotides, which are able to inhibit gene expression with exquisite specificity, are often used to elucidate the function of particular genes (see, for example, U.S. Pat. No. 6,165,790). Nucleic acid modulators are also used, 5 for example, to distinguish between functions of various members of a biological pathway. For example, antisense oligomers have been employed as therapeutic moieties in the treatment of disease states in animals and man and have been demonstrated in numerous clinical trials to be safe and effective (Milligan JF, *et al.*, Current Concepts in Antisense Drug Design, *J Med Chem.* (1993) 36:1923-1937; Tonkinson JL, *et al.*, Antisense 10 Oligodeoxynucleotides as Clinical Therapeutic Agents, *Cancer Invest.* (1996) 14:54-65). Accordingly, in one aspect of the invention, a MARK-specific nucleic acid modulator is used in an assay to further elucidate the role of the MARK in the p53 pathway, and/or its relationship to other members of the pathway. In another aspect of the invention, a 15 MARK-specific antisense oligomer is used as a therapeutic agent for treatment of p53-related disease states.

Assay Systems

The invention provides assay systems and screening methods for identifying specific modulators of MARK activity. As used herein, an "assay system" encompasses all the 20 components required for performing and analyzing results of an assay that detects and/or measures a particular event. In general, primary assays are used to identify or confirm a modulator's specific biochemical or molecular effect with respect to the MARK nucleic acid or protein. In general, secondary assays further assess the activity of a MARK modulating agent identified by a primary assay and may confirm that the modulating agent 25 affects MARK in a manner relevant to the p53 pathway. In some cases, MARK modulators will be directly tested in a secondary assay.

In a preferred embodiment, the screening method comprises contacting a suitable assay system comprising a MARK polypeptide with a candidate agent under conditions whereby, but for the presence of the agent, the system provides a reference activity (e.g., 30 kinase activity), which is based on the particular molecular event the screening method detects. A statistically significant difference between the agent-biased activity and the reference activity indicates that the candidate agent modulates MARK activity, and hence the p53 pathway.

Primary Assays

The type of modulator tested generally determines the type of primary assay.

Primary assays for small molecule modulators

For small molecule modulators, screening assays are used to identify candidate modulators. Screening assays may be cell-based or may use a cell-free system that recreates or retains the relevant biochemical reaction of the target protein (reviewed in Sittampalam GS *et al.*, Curr Opin Chem Biol (1997) 1:384-91 and accompanying references). As used herein the term "cell-based" refers to assays using live cells, dead cells, or a particular cellular fraction, such as a membrane, endoplasmic reticulum, or mitochondrial fraction. The term "cell free" encompasses assays using substantially purified protein (either endogenous or recombinantly produced), partially purified or crude cellular extracts. Screening assays may detect a variety of molecular events, including protein-DNA interactions, protein-protein interactions (*e.g.*, receptor-ligand binding), transcriptional activity (*e.g.*, using a reporter gene), enzymatic activity (*e.g.*, via a property of the substrate), activity of second messengers, immunogenicity and changes in cellular morphology or other cellular characteristics. Appropriate screening assays may use a wide range of detection methods including fluorescent, radioactive, colorimetric, spectrophotometric, and amperometric methods, to provide a read-out for the particular molecular event detected.

Cell-based screening assays usually require systems for recombinant expression of MARK and any auxiliary proteins demanded by the particular assay. Appropriate methods for generating recombinant proteins produce sufficient quantities of proteins that retain their relevant biological activities and are of sufficient purity to optimize activity and assure assay reproducibility. Yeast two-hybrid and variant screens, and mass spectrometry provide preferred methods for determining protein-protein interactions and elucidation of protein complexes. In certain applications, when MARK-interacting proteins are used in screens to identify small molecule modulators, the binding specificity of the interacting protein to the MARK protein may be assayed by various known methods such as substrate processing (*e.g.* ability of the candidate MARK-specific binding agents to function as negative effectors in MARK-expressing cells), binding equilibrium constants (usually at least about 10^7 M^{-1} , preferably at least about 10^8 M^{-1} , more preferably at least about 10^9 M^{-1}), and immunogenicity (*e.g.* ability to elicit MARK specific antibody

in a heterologous host such as a mouse, rat, goat or rabbit). For enzymes and receptors, binding may be assayed by, respectively, substrate and ligand processing.

The screening assay may measure a candidate agent's ability to specifically bind to or modulate activity of a MARK polypeptide, a fusion protein thereof, or to cells or membranes bearing the polypeptide or fusion protein. The MARK polypeptide can be full length or a fragment thereof that retains functional MARK activity. The MARK polypeptide may be fused to another polypeptide, such as a peptide tag for detection or anchoring, or to another tag. The MARK polypeptide is preferably human MARK, or is an ortholog or derivative thereof as described above. In a preferred embodiment, the screening assay detects candidate agent-based modulation of MARK interaction with a binding target, such as an endogenous or exogenous protein or other substrate that has MARK-specific binding activity, and can be used to assess normal MARK gene function.

Suitable assay formats that may be adapted to screen for MARK modulators are known in the art. Preferred screening assays are high throughput or ultra high throughput and thus provide automated, cost-effective means of screening compound libraries for lead compounds (Fernandes PB, Curr Opin Chem Biol (1998) 2:597-603; Sundberg SA, Curr Opin Biotechnol 2000, 11:47-53). In one preferred embodiment, screening assays uses fluorescence technologies, including fluorescence polarization, time-resolved fluorescence, and fluorescence resonance energy transfer. These systems offer means to monitor protein-protein or DNA-protein interactions in which the intensity of the signal emitted from dye-labeled molecules depends upon their interactions with partner molecules (e.g., Selvin PR, Nat Struct Biol (2000) 7:730-4; Fernandes PB, *supra*; Hertzberg RP and Pope AJ, Curr Opin Chem Biol (2000) 4:445-451).

A variety of suitable assay systems may be used to identify candidate MARK and p53 pathway modulators (e.g. U.S. Pat. No. 6,165,992 (kinase assays); U.S. Pat. Nos. 5,550,019 and 6,133,437 (apoptosis assays); U.S. Pat. No. 6,020,135 (p53 modulation), among others). Specific preferred assays are described in more detail below.

Kinase assays. In some preferred embodiments the screening assay detects the ability of the test agent to modulate the kinase activity of a MARK polypeptide. In further embodiments, a cell-free kinase assay system is used to identify a candidate p53 modulating agent, and a secondary, cell-based assay, such as an apoptosis or hypoxic induction assay (described below), may be used to further characterize the candidate p53 modulating agent. Many different assays for kinases have been reported in the literature

and are well known to those skilled in the art (e.g. U.S. Pat. No. 6,165,992; Zhu et al., Nature Genetics (2000) 26:283-289; and WO0073469). Radioassays, which monitor the transfer of a gamma phosphate are frequently used. For instance, a scintillation assay for p56 (lck) kinase activity monitors the transfer of the gamma phosphate from gamma-³³P ATP to a biotinylated peptide substrate; the substrate is captured on a streptavidin coated bead that transmits the signal (Beveridge M *et al.*, J Biomol Screen (2000) 5:205-212). This assay uses the scintillation proximity assay (SPA), in which only radio-ligand bound to receptors tethered to the surface of an SPA bead are detected by the scintillant immobilized within it, allowing binding to be measured without separation of bound from free ligand.

Other assays for protein kinase activity may use antibodies that specifically recognize phosphorylated substrates. For instance, the kinase receptor activation (KIRA) assay measures receptor tyrosine kinase activity by ligand stimulating the intact receptor in cultured cells, then capturing solubilized receptor with specific antibodies and quantifying phosphorylation via phosphotyrosine ELISA (Sadick MD, Dev Biol Stand (1999) 97:121-133).

Another example of antibody based assays for protein kinase activity is TRF (time-resolved fluorometry). This method utilizes europium chelate-labeled anti-phosphotyrosine antibodies to detect phosphate transfer to a polymeric substrate coated onto microtiter plate wells. The amount of phosphorylation is then detected using time-resolved, dissociation-enhanced fluorescence (Braunwalder AF, *et al.*, Anal Biochem 1996 Jul 1;238(2):159-64).

Apoptosis assays. Assays for apoptosis may be performed by terminal deoxynucleotidyl transferase-mediated digoxigenin-11-dUTP nick end labeling (TUNEL) assay. The TUNEL assay is used to measure nuclear DNA fragmentation characteristic of apoptosis (Lazebnik *et al.*, 1994, Nature 371, 346), by following the incorporation of fluorescein-dUTP (Yonehara *et al.*, 1989, J. Exp. Med. 169, 1747). Apoptosis may further be assayed by acridine orange staining of tissue culture cells (Lucas, R., *et al.*, 1998, Blood 15:4730-41). An apoptosis assay system may comprise a cell that expresses a MARK, and that optionally has defective p53 function (e.g. p53 is over-expressed or under-expressed relative to wild-type cells). A test agent can be added to the apoptosis assay system and changes in induction of apoptosis relative to controls where no test agent is added, identify candidate p53 modulating agents. In some embodiments of the invention, an apoptosis

assay may be used as a secondary assay to test a candidate p53 modulating agents that is initially identified using a cell-free assay system. An apoptosis assay may also be used to test whether MARK function plays a direct role in apoptosis. For example, an apoptosis assay may be performed on cells that over- or under-express MARK relative to wild type cells. Differences in apoptotic response compared to wild type cells suggests that the MARK plays a direct role in the apoptotic response. Apoptosis assays are described further in US Pat. No. 6,133,437.

Cell proliferation and cell cycle assays. Cell proliferation may be assayed via bromodeoxyuridine (BRDU) incorporation. This assay identifies a cell population undergoing DNA synthesis by incorporation of BRDU into newly-synthesized DNA. Newly-synthesized DNA may then be detected using an anti-BRDU antibody (Hoshino *et al.*, 1986, Int. J. Cancer 38, 369; Campana *et al.*, 1988, J. Immunol. Meth. 107, 79), or by other means.

Cell Proliferation may also be examined using [³H]-thymidine incorporation (Chen, J., 1996, Oncogene 13:1395-403; Jeoung, J., 1995, J. Biol. Chem. 270:18367-73). This assay allows for quantitative characterization of S-phase DNA syntheses. In this assay, cells synthesizing DNA will incorporate [³H]-thymidine into newly synthesized DNA. Incorporation can then be measured by standard techniques such as by counting of radioisotope in a scintillation counter (e.g., Beckman LS 3800 Liquid Scintillation Counter).

Cell proliferation may also be assayed by colony formation in soft agar (Sambrook *et al.*, Molecular Cloning, Cold Spring Harbor (1989)). For example, cells transformed with MARK are seeded in soft agar plates, and colonies are measured and counted after two weeks incubation.

Involvement of a gene in the cell cycle may be assayed by flow cytometry (Gray JW *et al.* (1986) Int J Radiat Biol Relat Stud Phys Chem Med 49:237-55). Cells transfected with a MARK may be stained with propidium iodide and evaluated in a flow cytometer (available from Becton Dickinson).

Accordingly, a cell proliferation or cell cycle assay system may comprise a cell that expresses a MARK, and that optionally has defective p53 function (e.g. p53 is over-expressed or under-expressed relative to wild-type cells). A test agent can be added to the assay system and changes in cell proliferation or cell cycle relative to controls where no test agent is added, identify candidate p53 modulating agents. In some embodiments of

the invention, the cell proliferation or cell cycle assay may be used as a secondary assay to test a candidate p53 modulating agents that is initially identified using another assay system such as a cell-free kinase assay system. A cell proliferation assay may also be used to test whether MARK function plays a direct role in cell proliferation or cell cycle. For example, a cell proliferation or cell cycle assay may be performed on cells that over- or under-express MARK relative to wild type cells. Differences in proliferation or cell cycle compared to wild type cells suggests that the MARK plays a direct role in cell proliferation or cell cycle.

Angiogenesis. Angiogenesis may be assayed using various human endothelial cell systems, such as umbilical vein, coronary artery, or dermal cells. Suitable assays include Alamar Blue based assays (available from Biosource International) to measure proliferation; migration assays using fluorescent molecules, such as the use of Becton Dickinson Falcon HTS FluoroBlock cell culture inserts to measure migration of cells through membranes in presence or absence of angiogenesis enhancer or suppressors; and tubule formation assays based on the formation of tubular structures by endothelial cells on Matrigel® (Becton Dickinson). Accordingly, an angiogenesis assay system may comprise a cell that expresses a MARK, and that optionally has defective p53 function (e.g. p53 is over-expressed or under-expressed relative to wild-type cells). A test agent can be added to the angiogenesis assay system and changes in angiogenesis relative to controls where no test agent is added, identify candidate p53 modulating agents. In some embodiments of the invention, the angiogenesis assay may be used as a secondary assay to test a candidate p53 modulating agents that is initially identified using another assay system. An angiogenesis assay may also be used to test whether MARK function plays a direct role in cell proliferation. For example, an angiogenesis assay may be performed on cells that over- or under-express MARK relative to wild type cells. Differences in angiogenesis compared to wild type cells suggests that the MARK plays a direct role in angiogenesis.

Hypoxic induction. The alpha subunit of the transcription factor, hypoxia inducible factor-1 (HIF-1), is upregulated in tumor cells following exposure to hypoxia in vitro. Under hypoxic conditions, HIF-1 stimulates the expression of genes known to be important in tumour cell survival, such as those encoding glyolytic enzymes and VEGF. Induction of such genes by hypoxic conditions may be assayed by growing cells

transfected with MARK in hypoxic conditions (such as with 0.1% O₂, 5% CO₂, and balance N₂, generated in a Napco 7001 incubator (Precision Scientific)) and normoxic conditions, followed by assessment of gene activity or expression by Taqman®. For example, a hypoxic induction assay system may comprise a cell that expresses a MARK, and that optionally has a mutated p53 (e.g. p53 is over-expressed or under-expressed relative to wild-type cells). A test agent can be added to the hypoxic induction assay system and changes in hypoxic response relative to controls where no test agent is added, identify candidate p53 modulating agents. In some embodiments of the invention, the hypoxic induction assay may be used as a secondary assay to test a candidate p53 modulating agents that is initially identified using another assay system. A hypoxic induction assay may also be used to test whether MARK function plays a direct role in the hypoxic response. For example, a hypoxic induction assay may be performed on cells that over- or under-express MARK relative to wild type cells. Differences in hypoxic response compared to wild type cells suggests that the MARK plays a direct role in hypoxic induction.

Cell adhesion. Cell adhesion assays measure adhesion of cells to purified adhesion proteins, or adhesion of cells to each other, in presence or absence of candidate modulating agents. Cell-protein adhesion assays measure the ability of agents to modulate the adhesion of cells to purified proteins. For example, recombinant proteins are produced, diluted to 2.5g/mL in PBS, and used to coat the wells of a microtiter plate. The wells used for negative control are not coated. Coated wells are then washed, blocked with 1% BSA, and washed again. Compounds are diluted to 2× final test concentration and added to the blocked, coated wells. Cells are then added to the wells, and the unbound cells are washed off. Retained cells are labeled directly on the plate by adding a membrane-permeable fluorescent dye, such as calcein-AM, and the signal is quantified in a fluorescent microplate reader.

Cell-cell adhesion assays measure the ability of agents to modulate binding of cell adhesion proteins with their native ligands. These assays use cells that naturally or recombinantly express the adhesion protein of choice. In an exemplary assay, cells expressing the cell adhesion protein are plated in wells of a multiwell plate. Cells expressing the ligand are labeled with a membrane-permeable fluorescent dye, such as BCECF, and allowed to adhere to the monolayers in the presence of candidate agents.

Unbound cells are washed off, and bound cells are detected using a fluorescence plate reader.

High-throughput cell adhesion assays have also been described. In one such assay, small molecule ligands and peptides are bound to the surface of microscope slides using a 5 microarray spotter, intact cells are then contacted with the slides, and unbound cells are washed off. In this assay, not only the binding specificity of the peptides and modulators against cell lines are determined, but also the functional cell signaling of attached cells using immunofluorescence techniques *in situ* on the microchip is measured (Falsey JR et al., Bioconjug Chem. 2001 May-Jun;12(3):346-53).

10

Primary assays for antibody modulators

For antibody modulators, appropriate primary assays test is a binding assay that tests the antibody's affinity to and specificity for the MARK protein. Methods for testing antibody affinity and specificity are well known in the art (Harlow and Lane, 1988, 1999, 15 *supra*). The enzyme-linked immunosorbant assay (ELISA) is a preferred method for detecting MARK-specific antibodies; others include FACS assays, radioimmunoassays, and fluorescent assays.

Primary assays for nucleic acid modulators

20 For nucleic acid modulators, primary assays may test the ability of the nucleic acid modulator to inhibit or enhance MARK gene expression, preferably mRNA expression. In general, expression analysis comprises comparing MARK expression in like populations of cells (*e.g.*, two pools of cells that endogenously or recombinantly express MARK) in the presence and absence of the nucleic acid modulator. Methods for analyzing mRNA 25 and protein expression are well known in the art. For instance, Northern blotting, slot blotting, ribonuclease protection, quantitative RT-PCR (*e.g.*, using the TaqMan®, PE Applied Biosystems), or microarray analysis may be used to confirm that MARK mRNA expression is reduced in cells treated with the nucleic acid modulator (*e.g.*, Current Protocols in Molecular Biology (1994) Ausubel FM *et al.*, *eds.*, John Wiley & Sons, Inc., 30 chapter 4; Freeman WM *et al.*, Biotechniques (1999) 26:112-125; Kallioniemi OP, Ann Med 2001, 33:142-147; Blohm DH and Guiseppi-Elie, A Curr Opin Biotechnol 2001, 12:41-47). Protein expression may also be monitored. Proteins are most commonly detected with specific antibodies or antisera directed against either the MARK protein or

specific peptides. A variety of means including Western blotting, ELISA, or in situ detection, are available (Harlow E and Lane D, 1988 and 1999, *supra*).

Secondary Assays

5 Secondary assays may be used to further assess the activity of MARK-modulating agent identified by any of the above methods to confirm that the modulating agent affects MARK in a manner relevant to the p53 pathway. As used herein, MARK-modulating agents encompass candidate clinical compounds or other agents derived from previously identified modulating agent. Secondary assays can also be used to test the activity of a
10 modulating agent on a particular genetic or biochemical pathway or to test the specificity of the modulating agent's interaction with MARK.

Secondary assays generally compare like populations of cells or animals (e.g., two pools of cells or animals that endogenously or recombinantly express MARK) in the presence and absence of the candidate modulator. In general, such assays test whether
15 treatment of cells or animals with a candidate MARK-modulating agent results in changes in the p53 pathway in comparison to untreated (or mock- or placebo-treated) cells or animals. Certain assays use "sensitized genetic backgrounds", which, as used herein, describe cells or animals engineered for altered expression of genes in the p53 or interacting pathways.

20

Cell-based assays

Cell based assays may use a variety of mammalian cell lines known to have defective p53 function (e.g. SAOS-2 osteoblasts, H1299 lung cancer cells, C33A and HT3 cervical cancer cells, HT-29 and DLD-1 colon cancer cells, among others, available from
25 American Type Culture Collection (ATCC), Manassas, VA). Cell based assays may detect endogenous p53 pathway activity or may rely on recombinant expression of p53 pathway components. Any of the aforementioned assays may be used in this cell-based format. Candidate modulators are typically added to the cell media but may also be injected into cells or delivered by any other efficacious means.

30

Animal Assays

A variety of non-human animal models of normal or defective p53 pathway may be used to test candidate MARK modulators. Models for defective p53 pathway typically use genetically modified animals that have been engineered to mis-express (e.g., over-express

or lack expression in) genes involved in the p53 pathway. Assays generally require systemic delivery of the candidate modulators, such as by oral administration, injection, etc.

In a preferred embodiment, p53 pathway activity is assessed by monitoring
5 neovascularization and angiogenesis. Animal models with defective and normal p53 are used to test the candidate modulator's affect on MARK in Matrigel® assays. Matrigel® is an extract of basement membrane proteins, and is composed primarily of laminin, collagen IV, and heparin sulfate proteoglycan. It is provided as a sterile liquid at 4° C, but rapidly forms a solid gel at 37° C. Liquid Matrigel® is mixed with various angiogenic agents,
10 such as bFGF and VEGF, or with human tumor cells which over-express the MARK. The mixture is then injected subcutaneously(SC) into female athymic nude mice (Taconic, Germantown, NY) to support an intense vascular response. Mice with Matrigel® pellets may be dosed via oral (PO), intraperitoneal (IP), or intravenous (IV) routes with the candidate modulator. Mice are euthanized 5 - 12 days post-injection, and the Matrigel®
15 pellet is harvested for hemoglobin analysis (Sigma plasma hemoglobin kit). Hemoglobin content of the gel is found to correlate the degree of neovascularization in the gel.

In another preferred embodiment, the effect of the candidate modulator on MARK is assessed via tumorigenicity assays. In one example, xenograft human tumors are implanted SC into female athymic mice, 6-7 week old, as single cell suspensions either
20 from a pre-existing tumor or from *in vitro* culture. The tumors which express the MARK endogenously are injected in the flank, 1 x 10⁵ to 1 x 10⁷ cells per mouse in a volume of 100 µL using a 27gauge needle. Mice are then ear tagged and tumors are measured twice weekly. Candidate modulator treatment is initiated on the day the mean tumor weight reaches 100 mg. Candidate modulator is delivered IV, SC, IP, or PO by bolus
25 administration. Depending upon the pharmacokinetics of each unique candidate modulator, dosing can be performed multiple times per day. The tumor weight is assessed by measuring perpendicular diameters with a caliper and calculated by multiplying the measurements of diameters in two dimensions. At the end of the experiment, the excised tumors maybe utilized for biomarker identification or further analyses. For
30 immunohistochemistry staining, xenograft tumors are fixed in 4% paraformaldehyde, 0.1M phosphate, pH 7.2, for 6 hours at 4° C, immersed in 30% sucrose in PBS, and rapidly frozen in isopentane cooled with liquid nitrogen.

Diagnostic and therapeutic uses

Specific MARK-modulating agents are useful in a variety of diagnostic and therapeutic applications where disease or disease prognosis is related to defects in the p53 pathway, such as angiogenic, apoptotic, or cell proliferation disorders. Accordingly, the invention also provides methods for modulating the p53 pathway in a cell, preferably a cell pre-determined to have defective p53 function, comprising the step of administering an agent to the cell that specifically modulates MARK activity. Preferably, the modulating agent produces a detectable phenotypic change in the cell indicating that the p53 function is restored, i.e., for example, the cell undergoes normal proliferation or progression through the cell cycle.

The discovery that MARK is implicated in p53 pathway provides for a variety of methods that can be employed for the diagnostic and prognostic evaluation of diseases and disorders involving defects in the p53 pathway and for the identification of subjects having a predisposition to such diseases and disorders.

Various expression analysis methods can be used to diagnose whether MARK expression occurs in a particular sample, including Northern blotting, slot blotting, ribonuclease protection, quantitative RT-PCR, and microarray analysis. (e.g., Current Protocols in Molecular Biology (1994) Ausubel FM *et al.*, eds., John Wiley & Sons, Inc., chapter 4; Freeman WM *et al.*, Biotechniques (1999) 26:112-125; Kallioniemi OP, Ann Med 2001, 33:142-147; Blohm and Guiseppi-Elie, Curr Opin Biotechnol 2001, 12:41-47). Tissues having a disease or disorder implicating defective p53 signaling that express a MARK, are identified as amenable to treatment with a MARK modulating agent. In a preferred application, the p53 defective tissue overexpresses a MARK relative to normal tissue. For example, a Northern blot analysis of mRNA from tumor and normal cell lines, or from tumor and matching normal tissue samples from the same patient, using full or partial MARK cDNA sequences as probes, can determine whether particular tumors express or overexpress MARK. Alternatively, the TaqMan® is used for quantitative RT-PCR analysis of MARK expression in cell lines, normal tissues and tumor samples (PE Applied Biosystems).

Various other diagnostic methods may be performed, for example, utilizing reagents such as the MARK oligonucleotides, and antibodies directed against a MARK, as described above for: (1) the detection of the presence of MARK gene mutations, or the detection of either over- or under-expression of MARK mRNA relative to the non-disorder state; (2) the detection of either an over- or an under-abundance of MARK gene

product relative to the non-disorder state; and (3) the detection of perturbations or abnormalities in the signal transduction pathway mediated by MARK.

Thus, in a specific embodiment, the invention is drawn to a method for diagnosing a disease in a patient, the method comprising: a) obtaining a biological sample from the 5 patient; b) contacting the sample with a probe for MARK expression; c) comparing results from step (b) with a control; and d) determining whether step (c) indicates a likelihood of disease. Preferably, the disease is cancer, most preferably a cancer as shown in TABLE 1. The probe may be either DNA or protein, including an antibody.

10 EXAMPLES

The following experimental section and examples are offered by way of illustration and not by way of limitation.

I. Drosophila p53 screen

15 The Drosophila p53 gene was overexpressed specifically in the wing using the vestigial margin quadrant enhancer. Increasing quantities of Drosophila p53 (titrated using different strength transgenic inserts in 1 or 2 copies) caused deterioration of normal wing morphology from mild to strong, with phenotypes including disruption of pattern and polarity of wing hairs, shortening and thickening of wing veins, progressive crumpling of 20 the wing and appearance of dark "death" inclusions in wing blade. In a screen designed to identify enhancers and suppressors of Drosophila p53, homozygous females carrying two copies of p53 were crossed to 5663 males carrying random insertions of a piggyBac transposon (Fraser M *et al.*, Virology (1985) 145:356-361). Progeny containing insertions 25 were compared to non-insertion-bearing sibling progeny for enhancement or suppression of the p53 phenotypes. Sequence information surrounding the piggyBac insertion site was used to identify the modifier genes. Modifiers of the wing phenotype were identified as members of the p53 pathway. kp78a was a suppressor of the wing phenotype. Human orthologs of the modifiers are referred to herein as MARK.

BLAST analysis (Altschul *et al.*, *supra*) was employed to identify Targets from 30 Drosophila modifiers. For example, representative sequences from MARK GI# 9845487 (SEQ ID NO:24), GI# 8923922(SEQ ID NO:25), GI# 4505103 (SEQ ID NO:27), and GI#13899225 (SEQ ID NO:29) share 43%, 65%, 65% and 45% amino acid identity, respectively, with the *Drosophila* kp78a.

Various domains, signals, and functional subunits in proteins were analyzed using the PSORT (Nakai K., and Horton P., Trends Biochem Sci, 1999, 24:34-6; Kenta Nakai, Protein sorting signals and prediction of subcellular localization, Adv. Protein Chem. 54, 277-344 (2000)), PFAM (Bateman A., et al., Nucleic Acids Res, 1999, 27:260-2;

5 <http://pfam.wustl.edu>), SMART (Ponting CP, et al., SMART: identification and annotation of domains from signaling and extracellular protein sequences. Nucleic Acids Res. 1999 Jan 1;27(1):229-32), TM-HMM (Erik L.L. Sonnhammer, Gunnar von Heijne, and Anders Krogh: A hidden Markov model for predicting transmembrane helices in protein sequences. In Proc. of Sixth Int. Conf. on Intelligent Systems for Molecular Biology, p 10 175-182 Ed J. Glasgow, T. Littlejohn, F. Major, R. Lathrop, D. Sankoff, and C. Sensen Menlo Park, CA: AAAI Press, 1998), and clust (Remm M, and Sonnhammer E. Classification of transmembrane protein families in the *Caenorhabditis elegans* genome and identification of human orthologs. Genome Res. 2000 Nov;10(11):1679-89) programs. For example, the protein kinase domains of MARKs from GI#s 9845487 (SEQ ID NO:24), 15 8923922 (SEQ ID NO:25), 4505103 (SEQ ID NO:27), and 13899225 (SEQ ID NO:29) is located at approximately amino acid residues 20 to 271, 60 to 311, 56 to 307, and 59 to 310, respectively (PFAM 00069). Further, the ubiquitin associated (UBA/TS-N) domains of MARKs from GI#s 9845487 (SEQ ID NO:24), 8923922 (SEQ ID NO:25), 4505103 (SEQ ID NO:27), and 13899225 (SEQ ID NO:29) is located at approximately amino acid residues 291 to 330, 331 to 370, 327 to 366, and 330 to 369, respectively (PFAM 00627). Still further, the kinase associated domains from MARKs of GI#s 9845487 (SEQ ID NO:24), 8923922 (SEQ ID NO:25), and 4505103 (SEQ ID NO:27) are located at approximately amino acid residues 696 to 745, 746 to 795, and 664 to 713, respectively (PFAM 02149).

25

II. High-Throughput In Vitro Fluorescence Polarization Assay

Fluorescently-labeled MARK peptide/substrate are added to each well of a 96-well microtiter plate, along with a test agent in a test buffer (10 mM HEPES, 10 mM NaCl, 6 mM magnesium chloride, pH 7.6). Changes in fluorescence polarization, determined by 30 using a Fluorolite FPM-2 Fluorescence Polarization Microtiter System (Dynatech Laboratories, Inc), relative to control values indicates the test compound is a candidate modifier of MARK activity.

III. High-Throughput In Vitro Binding Assay.

³³P-labeled MARK peptide is added in an assay buffer (100 mM KCl, 20 mM HEPES pH 7.6, 1 mM MgCl₂, 1% glycerol, 0.5% NP-40, 50 mM beta-mercaptoethanol, 1 mg/ml BSA, cocktail of protease inhibitors) along with a test agent to the wells of a Neutralite-5 avidin coated assay plate and incubated at 25°C for 1 hour. Biotinylated substrate is then added to each well and incubated for 1 hour. Reactions are stopped by washing with PBS, and counted in a scintillation counter. Test agents that cause a difference in activity relative to control without test agent are identified as candidate p53 modulating agents.

10 IV. Immunoprecipitations and Immunoblotting

For coprecipitation of transfected proteins, 3 × 10⁶ appropriate recombinant cells containing the MARK proteins are plated on 10-cm dishes and transfected on the following day with expression constructs. The total amount of DNA is kept constant in each transfection by adding empty vector. After 24 h, cells are collected, washed once 15 with phosphate-buffered saline and lysed for 20 min on ice in 1 ml of lysis buffer containing 50 mM Hepes, pH 7.9, 250 mM NaCl, 20 mM -glycerophosphate, 1 mM sodium orthovanadate, 5 mM p-nitrophenyl phosphate, 2 mM dithiothreitol, protease inhibitors (complete, Roche Molecular Biochemicals), and 1% Nonidet P-40. Cellular debris is removed by centrifugation twice at 15,000 × g for 15 min. The cell lysate is 20 incubated with 25 µl of M2 beads (Sigma) for 2 h at 4 °C with gentle rocking.

After extensive washing with lysis buffer, proteins bound to the beads are solubilized by boiling in SDS sample buffer, fractionated by SDS-polyacrylamide gel electrophoresis, transferred to polyvinylidene difluoride membrane and blotted with the indicated antibodies. The reactive bands are visualized with horseradish peroxidase coupled to the 25 appropriate secondary antibodies and the enhanced chemiluminescence (ECL) Western blotting detection system (Amersham Pharmacia Biotech).

V. Kinase assay

A purified or partially purified MARK is diluted in a suitable reaction buffer, e.g., 50 30 mM Hepes, pH 7.5, containing magnesium chloride or manganese chloride (1-20 mM) and a peptide or polypeptide substrate, such as myelin basic protein or casein (1-10 µg/ml). The final concentration of the kinase is 1-20 nM. The enzyme reaction is conducted in microtiter plates to facilitate optimization of reaction conditions by increasing assay throughput. A 96-well microtiter plate is employed using a final volume 30-100 µl. The

reaction is initiated by the addition of ^{33}P -gamma-ATP (0.5 $\mu\text{Ci}/\text{ml}$) and incubated for 0.5 to 3 hours at room temperature. Negative controls are provided by the addition of EDTA, which chelates the divalent cation (Mg^{2+} or Mn^{2+}) required for enzymatic activity.

Following the incubation, the enzyme reaction is quenched using EDTA. Samples of the
5 reaction are transferred to a 96-well glass fiber filter plate (MultiScreen, Millipore). The filters are subsequently washed with phosphate-buffered saline, dilute phosphoric acid (0.5%) or other suitable medium to remove excess radiolabeled ATP. Scintillation cocktail is added to the filter plate and the incorporated radioactivity is quantitated by scintillation counting (Wallac/Perkin Elmer). Activity is defined by the amount of
10 radioactivity detected following subtraction of the negative control reaction value (EDTA quench).

VI. Expression analysis

All cell lines used in the following experiments are NCI (National Cancer Institute)
15 lines, and are available from ATCC (American Type Culture Collection, Manassas, VA 20110-2209). Normal and tumor tissues were obtained from Impath, UC Davis, Clontech, Stratagene, and Ambion.

TaqMan analysis was used to assess expression levels of the disclosed genes in various samples.
20 RNA was extracted from each tissue sample using Qiagen (Valencia, CA) RNeasy kits, following manufacturer's protocols, to a final concentration of 50ng/ μl . Single stranded cDNA was then synthesized by reverse transcribing the RNA samples using random hexamers and 500ng of total RNA per reaction, following protocol 4304965 of Applied Biosystems (Foster City, CA, <http://www.appliedbiosystems.com/>).

25 Primers for expression analysis using TaqMan assay (Applied Biosystems, Foster City, CA) were prepared according to the TaqMan protocols, and the following criteria: a) primer pairs were designed to span introns to eliminate genomic contamination, and b) each primer pair produced only one product.

Taqman reactions were carried out following manufacturer's protocols, in 25 μl total
30 volume for 96-well plates and 10 μl total volume for 384-well plates, using 300nM primer and 250 nM probe, and approximately 25ng of cDNA. The standard curve for result analysis was prepared using a universal pool of human cDNA samples, which is a mixture of cDNAs from a wide variety of tissues so that the chance that a target will be present in

appreciable amounts is good. The raw data were normalized using 18S rRNA (universally expressed in all tissues and cells).

For each expression analysis, tumor tissue samples were compared with matched normal tissues from the same patient. A gene was considered overexpressed in a tumor when the level of expression of the gene was 2 fold or higher in the tumor compared with its matched normal sample.

Results are shown in Table 1. Data presented in bold indicate that greater than 50% of tested tumor samples of the tissue type indicated in row 1 exhibited over expression of the gene listed in column 1, relative to normal samples. Underlined data indicates that between 25% to 49% of tested tumor samples exhibited over expression. A modulator identified by an assay described herein can be further validated for therapeutic effect by administration to a tumor in which the gene is overexpressed. A decrease in tumor growth confirms therapeutic utility of the modulator. Prior to treating a patient with the modulator, the likelihood that the patient will respond to treatment can be diagnosed by obtaining a tumor sample from the patient, and assaying for expression of the gene targeted by the modulator. The expression data for the gene(s) can also be used as a diagnostic marker for disease progression. The assay can be performed by expression analysis as described above, by antibody directed to the gene target, or by any other available detection method.

20

Table 1

	<u>breast</u>	<u>colon</u>	<u>lung</u>	<u>ovary</u>
GI#9845486 (SEQ ID NO:1)	7 11.	8 30.	8 13.	5 7
GI#9845488 (SEQ ID NO:2)	1 11.	4 30.	0 13.	1 7
GI#8923921(SEQ ID NO:8)	2 11.	7 30.	<u>6</u> <u>13</u> .	0 7
GI#3089348 (SEQ ID NO:13)	2 11.	2 30.	0 13.	<u>2</u> 7
GI#13366083 (SEQ ID NO:19)	2 11.	2 30.	<u>5</u> <u>13</u> .	1 7

WHAT IS CLAIMED IS:

1. A method of identifying a candidate p53 pathway modulating agent, said method comprising the steps of:
 - 5 (a) providing an assay system comprising a purified MARK polypeptide or nucleic acid or a functionally active fragment or derivative thereof;
 - (b) contacting the assay system with a test agent under conditions whereby, but for the presence of the test agent, the system provides a reference activity; and
 - (c) detecting a test agent-biased activity of the assay system, wherein a difference
- 10 between the test agent-biased activity and the reference activity identifies the test agent as a candidate p53 pathway modulating agent.
2. The method of Claim 1 wherein the assay system comprises cultured cells that express the MARK polypeptide.
- 15 3. The method of Claim 2 wherein the cultured cells additionally have defective p53 function.
4. The method of Claim 1 wherein the assay system includes a screening assay
- 20 comprising a MARK polypeptide, and the candidate test agent is a small molecule modulator.
5. The method of Claim 4 wherein the assay is a kinase assay.
- 25 6. The method of Claim 1 wherein the assay system is selected from the group consisting of an apoptosis assay system, a cell proliferation assay system, an angiogenesis assay system, and a hypoxic induction assay system.
7. The method of Claim 1 wherein the assay system includes a binding assay
- 30 comprising a MARK polypeptide and the candidate test agent is an antibody.
8. The method of Claim 1 wherein the assay system includes an expression assay comprising a MARK nucleic acid and the candidate test agent is a nucleic acid modulator.

9. The method of claim 8 wherein the nucleic acid modulator is an antisense oligomer.
10. The method of Claim 8 wherein the nucleic acid modulator is a PMO.
5
11. The method of Claim 1 additionally comprising:
 - (d) administering the candidate p53 pathway modulating agent identified in (c) to a model system comprising cells defective in p53 function and, detecting a phenotypic change in the model system that indicates that the p53 function is restored.
- 10
12. The method of Claim 11 wherein the model system is a mouse model with defective p53 function.
13. A method for modulating a p53 pathway of a cell comprising contacting a cell
15 defective in p53 function with a candidate modulator that specifically binds to a MARK polypeptide comprising an amino acid sequence selected from group consisting of SEQ ID NOs:24, 25, 26, 27, 28, and 29, whereby p53 function is restored.
14. The method of claim 13 wherein the candidate modulator is administered to a
20 vertebrate animal predetermined to have a disease or disorder resulting from a defect in p53 function.
15. The method of Claim 13 wherein the candidate modulator is selected from the
group consisting of an antibody and a small molecule.
25
16. The method of Claim 1, comprising the additional steps of:
 - (d) providing a secondary assay system comprising cultured cells or a non-human animal expressing MARK ,
 - (e) contacting the secondary assay system with the test agent of (b) or an agent derived therefrom under conditions whereby, but for the presence of the test agent or agent derived therefrom, the system provides a reference activity; and
 - 30 (f) detecting an agent-biased activity of the second assay system,

wherein a difference between the agent-biased activity and the reference activity of the second assay system confirms the test agent or agent derived therefrom as a candidate p53 pathway modulating agent,
and wherein the second assay detects an agent-biased change in the p53 pathway.

5

17. The method of Claim 16 wherein the secondary assay system comprises cultured cells.

18. The method of Claim 16 wherein the secondary assay system comprises a non-
10 human animal.

19. The method of Claim 18 wherein the non-human animal mis-expresses a p53 pathway gene.

15 20. A method of modulating p53 pathway in a mammalian cell comprising contacting the cell with an agent that specifically binds a MARK polypeptide or nucleic acid.

21. The method of Claim 20 wherein the agent is administered to a mammalian animal predetermined to have a pathology associated with the p53 pathway.

20

22. The method of Claim 20 wherein the agent is a small molecule modulator, a nucleic acid modulator, or an antibody.

23. A method for diagnosing a disease in a patient comprising:
25 (a) obtaining a biological sample from the patient;
(b) contacting the sample with a probe for MARK expression;
(c) comparing results from step (b) with a control;
(d) determining whether step (c) indicates a likelihood of disease.

30 24. The method of claim 23 wherein said disease is cancer.

25. The method according to claim 24, wherein said cancer is a cancer as shown in Table 1 as having >25% expression level.

SEQUENCE LISTING

<110> EXELIXIS, INC.

<120> MARKS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE

<130> EX02-088C-PC

<150> US 60/296,076

<151> 2001-06-05

<150> US 60/328,605

<151> 2001-10-10

<150> US 60/357,253

<151> 2002-02-15

<160> 29

<170> PatentIn version 3.1

<210> 1

<211> 2946

<212> DNA

<213> Homo sapiens

<400> 1
tccctggaatt gcacgcgtt cctgaccacc aggctctggc ccttgagaag ccagcgggc 60
tttgcgtccctg ttgctctct tgccaaaccc agtctctctg ctatgtggtgg tttcggttgc 120
gacaccgtcc aggttcccag gcaggaacctg ctcggcctgg ctgcttagct acttttcaact 180
gaggaggtgg tggaaagggtgt cgccctgtct ggctgagtaa ggggtggctgg ctgagccggc 240
agccccccgcc ctaggcctgg ctcttcccgg cctctgtact ttgccttcgc tgcctgacag 300
gttctgttgt gggctctgtc gaatggaagt cgctggtagt cctttccct ttctccagtc 360
ggcccacctt gggacacccctt gactccaagc ccagcagtaa gtccaaacatg attcggggcc 420
gcaactcagc cacctctgtc gatgagcagc cccacattgg aaactaccgg ctcctcaaga 480
ccattggcaa gggtaatttt gccaaagggtga agttggcccg acacatcctg actgggaaag 540
aggtagctgt gaagatcatt gacaagactc aactgaactc ctccagcctc cagaaaactat 600
tccgcgaagt aagaataatg aagggtttga atcatccaa catagttaaa ttatttgaag 660
tgatttagac tgagaaaaacg ctctaccttg tcatggagta cgcttagtggc ggagaggtat 720
ttgattacct agtggctcat ggcaggatga aagaaaaaga ggctcgagcc aaattccgccc 780
agatagtgtc tgctgtgcag tactgtcacc agaagtttat tgtccataga gacttaaagg 840
cagaaaaacct gctcttggat gctgatatga acatcaagat tgcagacttt ggcttcagca 900
atgaattcac ctggaaac aagctggaca cttctgtgg cagtccttcat tatgctgccc 960
cagaactctt ccaggcataaa aaatatgtg gacccgaggt ggatgtgtgg agccttaggag 1020
ttatcctcta tacactggtc agcgatccc tgcctttga tggacagaac ctcaaggagc 1080

tgcccccaacg ggtactgagg gggaaatacc gtattccatt ctacatgtcc acggactgtg 1140
 aaaacctgct taagaaatcc ctcatcctta atcccagcaa gagaggcact tttagagcaaa 1200
 tcatgaaaga tcgatggatg aatgtgggtc acgaagatga tgaactaaag ccttacgtgg 1260
 agccactccc tgactacaag gaccccccgc ggacagagct gatgggtgtcc atgggttata 1320
 cacgggaaga gatccaggac tcgctgggtgg gccagagata caacgaggtg atggccacct 1380
 atctgctcct gggctacaag agctccgagc tggaaaggcg aaccatcacc ctgaaacccc 1440
 ggccttcagc tgatctaacc aatagcagcg cccaaattccc atcccacaag gtacagcgaa 1500
 gcgtgtcggc caatcccaag cagcggcgct tcagcgacca ggctggtcct gccattccca 1560
 cctctaattc ttactctaag aagactcaga gtaacaacgc agaaaataag cggcctgagg 1620
 aggaccggga gtcagggcgg aaagccagca gcacagccaa ggtgcctgtcc agccccctgc 1680
 ccggctctgga gaggaagaag accacccaa cccctccac gaacagcgcc ctatccacca 1740
 gcacaaaatcg aagcaggaat tccccacttt tggagcgggc cagcctcgcc caggcctcca 1800
 tccagaatgg caaagacagc ctaaccatgc cagggtccccg ggcctccacg gttctgttt 1860
 ctggccgagt ctctcgcccgcc cggcccccgc agcaccagaa atccatgtcg gcctccgtgc 1920
 accccaacaa ggcctctggg ctgccccccca cggagagtaa ctgtgagggtg ccgcggccca 1980
 gcacagcccc ccagcgtgtc cctgttgact ccccatccgc ccacaacatc agcagcagtg 2040
 gtggagcccc agaccgaact aacttccccc ggggtgtgtc cagccgaagc accttccatg 2100
 ctgggcagct ccgacaggtg cgggaccagc agaatttgcc ctacggtgtg accccagct 2160
 ctccctctgg ccacagccag ggcggcggg gggcctctgg gagcatcttc agcaagttca 2220
 cctccaaagtt tgtacgcagg aacctgaatg aacctgaaaag caaagaccga gtggagacgc 2280
 tcagaccta cgtgggtggc agtggcggca acgacaaaaga aaaggaagaa tticgggagg 2340
 ccaagccccg ctccctccgc ttacacgtggaa gtatgaagac cacgagctcc atggagccca 2400
 acgagatgtat gcgggagatc cgcaagggtgc tggacgcgaa cagctgccag agcagcgtgc 2460
 atgagaagta catgctgtgc tgcatgcacgc gcacggcggg ccacgaggac ttctgtgcagt 2520
 gggagatgga ggtgtcaaa ctgcccggc tctctctcaa cggggttcga tttaagcgga 2580
 tatcgggcac ctccatggcc ttcaaaaaca ttgcctccaa aatagccaaac gagctgaagc 2640
 tttaacaggc tgccaggagc gggggcggcgg gggggcggcc agctggacgg gctgcccggcc 2700
 gtgcggccgc ccacctggc gagactgcag cgatggattt gttgttatcc ctgctggcac 2760
 ttctcccttc ctggccctt ctcaaggtttc tcccacattc accccctgcggcc agagattccc 2820
 ccttcttcctc tccccactg gaggcaaagg aaggggaggg tggatggggg ggcagggtc 2880
 cccctcggttta ctgcgggtgc acagagtatt tcgcctaaac caagaaattt tttattacca 2940

aaaaga	2946
<210> 2	
<211> 2784	
<212> DNA	
<213> Homo sapiens	
<400> 2	
tcctgaaatt gcacgcgtt cctgaccacc aggatctggc ctttggagaag ccagcgcccc	60
tttgtccctg ttgtctctct tgccaaaccc agtctctctg ctatgtggtgg tttcggttgc	120
gacaccgtcc aggttcccag gcaggaacct ctggcctgg ctgcttagct acttttact	180
gaggaggtgg tggaaagggtgt cgccctgctct ggctgagtaa gggtggtgg ctgagccggc	240
agcccccgcc cttaggcctgg ctcttcccg cctctgtact ttgcctcgc tgcctgacag	300
gttctgctgt gggctctgct gaatggaagt cgctggtagt cttttccct ttctccagtc	360
ggcccacctt gggacacctt gactccaagc ccagcagtaa gtccaaacatg attcgcccc	420
gcaactcagc cacctctgct gatgagcagc cccacattgg aaactaccgg ctccctcaaga	480
ccattggcaa gggtaatttt gccaagggtga agttggcccg acacatcctg actggaaag	540
aggtagctgt gaagatcatt gacaagactc aactgaactc ctccagcctc cagaaactat	600
tccgcgaagt aagaataatg aaggtttga atcatccaa catagttaaa ttatgttgaag	660
tgattgagac tgagaaaacg ctctaccttg tcatggagta cgctagtgcc ggagaggtat	720
ttgattacct agtggctcat ggcaggatga aagaaaaaga ggctcgagcc aaattccgcc	780
agatagtgtc tgctgtgcag tactgtcacc agaagtttat tgtccataga gacttaaagg	840
cagaaaaacct gctcttggat gctgatatga acatcaagat tgcagacttt ggcttcagca	900
atgaattcac ctggaaac aagctggaca ctttctgtgg cagttccct tatgtgtccc	960
cagaactctt ccaggcddd aaatatgtg gacccgaggt ggatgtgtgg agcctaggag	1020
ttatctcta tacactggtc agcggatccc tgcctttga tggacagaac ctcaaggagc	1080
tgcggaaacg ggtactgagg gggaaatacc gtattccatt ctacatgtcc acggactgtg	1140
aaaacctgct taagaaaattt ctcatccta atcccagcaa gagaggact tttagagcaaa	1200
tcatgaaaaga tcgatggatg aatgtgggtc acgaagatga tgaactaaag ctttacgtgg	1260
agccactccc tgactacaag gaccccccggc ggacagagct gatggtgtcc atgggttata	1320
cacggaaaga gatccaggac tcgctggtgg gccagagata caacgaggtg atggccacct	1380
atctgtctctt gggctacaag agctccgagc tggaaaggcga caccatcacc ctgaaacccc	1440
ggccttcagc tgatctaacc aatagcagcg cccaaattccc atcccacaag gtacagcgaa	1500
ggctgtcggc caatcccaag cagccggcgt tcagcgacca ggctggtcct gccattccca	1560

cctctaattc ttactctaag aagactcaga gtaacaacgc agaaaataag cggcctgagg	1620
aggaccggga gtcagggcgg aaagccagca gcacagccaa ggtgcctgcc agccccctgc	1680
ccgggtctgga gaggaagaag accaccccaa cccccctccac gaacagcgtc ctctccacca	1740
gcacaaatcg aagcaggaat tcccccactt tggagcgggc cagcctcggc caggcctcca	1800
tccagaatgg caaagacagc acagccccc agcgtgtccc tttgcctcc ccatccgccc	1860
acaacatcg cagcagtggt ggagcccccag accgaactaa cttccccgg ggtgtgtcca	1920
gccgaagcac cttccatgct gggcagctcc gacaggtgcg ggaccagcag aatttgcct	1980
acggtgtgac cccagccctc ccctctggcc acagccaggg cggcgggggg gcctctggga	2040
gcacatccatg caagttcacc tccaagttt tacgcaggaa cctgaatgaa cctgaaagca	2100
aagaccgagt ggagacgctc agacctcaco tggtggcag tggcggcaac gacaaagaaa	2160
aggaagaatt tcgggaggcc aagccccgct ccctccgctt cacgtggagt atgaagacca	2220
cgagctccat ggagcccaac gagatgatgc gggagatccg caaggtgctg gacgcgaaca	2280
gctgccagag cgagctgcat gagaagtaca tgctgctgt catgcacggc acgcggggcc	2340
acgaggactt cgtgcagtgg gagatggagg tgtgcaaact gccgcggctc tctctcaacg	2400
gggttcgatt taagcggata tcgggcaccc ccatggcctt caaaaacatt gcctccaaaa	2460
tagccaacga gctgaagctt taacaggctg ccaggagcgg gggcggcggg ggcgggcccag	2520
ctggacgggc tgccggccgt gcccggcccc acctgggcga gactgcagcg atggatttgt	2580
gtgtctccct gctggcactt ctccccctcc tggcccttct cagttttctc ccacattcac	2640
ccctgcccag agattcccc ttctcctctc ccctactgga ggcaaaggaa ggggagggtg	2700
gatggggggg cagggctccc cctcggtact gcggttgcac agagtatttc gcctaaacca	2760
agaaaattttt tattaccaaa aaga	2784

<210> 3
 <211> 3103
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (2941)..(2941)
 <223> "n" is A, C, G, or T

cgggtggtggc ggccatgttg ggagcagcag gtccggcggc ggctgcctgt gtgcggcggc	60
cggagcagtg ccgctgaggg caggggagga gcgaggcagg cggccggctg cggcggcaga	120
gagttaggcgg agcggcgcgg cccggccgaa aggccgcaca gcccagccgg ggtcgggggg	180
ggtgccgtcc ggagccgctc ggagccggcg cggcctagcc cgagcggcgc atccccggc	240

tggcgtgagc ggctgccccgg cctcccccga cccccggccgg gggcccatgc ggccgggtgct	300
cctgctgtga gaagccccgc ccggccgggc tccgcgcatt cccttccctc ctttcctcca	360
agtttcgtcg ttccctcccc cgagataccg ggcgcatgtc cagcgctcgg accccccctac	420
ccacgctgaa cgagagggac acggagcagc ccaccttggg acaccttgac tccaagccca	480
gcagtaagtc caacatgatt cggggccgca actcagccac ctctgctgat gagcagcccc	540
acattggaaa ctaccggctc ctcaagacca ttggcaaggg taattttgcc aaggtgaagt	600
tggcccgaca catcctgact gggaaagagg tagctgtgaa gatcatttgac aagactcaac	660
tgaactcctc cagcctccag aaactattcc gcgaagtaag aataatgaag gttttgaatc	720
atcccaacat agttaaatatta tttgaagtga ttgagactga gaaaacgctc taccttgc	780
tggagtagc tagtggcgga gaggtatttg attacctagt ggctcatggc aggatgaaag	840
aaaaagaggc tcgagccaaa ttccggcaga tagtgtctgc tgtgcagtc tgtcaccaga	900
agtttattgt ccatagagac taaaaggcag aaaacctgct cttggatgct gatataaca	960
tcaagattgc agactttggc ttcagcaatg aattcacctt tggaaacaag ctggacac	1020
tctgtggcag tcccccttat gctgccccag aactcttcca gggcaaaaaaa tatgatggac	1080
ccgaggtgga tgtgtggagc ctaggagtta tcctctatac actggcagc ggatccctgc	1140
cttttgc acagaacctc aaggagctgc gggAACGGGT actgagggaa aaataccgt	1200
ttccattcta catgtccacg gactgtgaaa acctgcttaa gaaatttctc attcttaatc	1260
ccagcaagag aggcaccaa gagcaaatca tgaaagatcg atggatgaat gtgggtcacg	1320
aagatgatga actaaagcct tacgtggagc cactccctga ctacaaggac cccggcgga	1380
cagagctgat ggtgtccatg gtttatacac gggaaagagat ccaggactcg ctgggtggcc	1440
agagatacaa cgaggtgatg gccacctatac tgctcctggg ctacaaggac tccgagctgg	1500
aaggcgacac catcacccctg aaaccccgcc cttcagctga tctgaccaat agcagcgccc	1560
catccccatc ccacaaggta cagcgcagcg tgtcgccaa tcccaaggac cggcgcttca	1620
gcgaccaggc tggcctgcc attcccacct ctaattctta ctctaagaag actcagagta	1680
acaacgcaga aaataagcgg cctgaggagg accgggagtc agggcggaaa gccagcagca	1740
cagccaaagggt gctgtccagc cccctgcccc gtctggagag gaagaagacc accccaaccc	1800
cctccacggc acagcgtcct ctccaccaggc acaaatcgaa gcaggaattc cccacttttgc	1860
gagcggggca gcctcgccca ggcctccatc cagaatggca aagacagcct aaccatgcc	1920
gggtccccggg cctccacggc ttctgcttct gccgcagtct ctgcggccccg gccccggccag	1980
caccagaaat ccatgtcgcc ctccgtgcac cccaaacaagg cctctggct gccccccacg	2040
gagagtaact gtgaggtgcc gcggcccgac acagcccccc agcgtgtccc tggtgcctcc	2100

ccatccgccc acaacatcatcag cagcagtgggt ggagccccag accgaactaa cttcccccg	2160
ggtgtgtcca gccgaagcac cttccatgct gggcagctcc gacagggtgcg ggaccagcag	2220
aatttgcctt acggtgtgac cccagcctct ccctctggcc acagccaggg ccggcggggg	2280
gcctctggga gcacatccatcag caagttcacc tccaaagtgg tacgcagaaa tctgtcttc	2340
agggttgcca gaaggaacct gaatgaacct gaaagcaaag accgagtgga gacgctcaga	2400
cctcacgtgg tgggcagtgg cgcaacgcgaa aaagaaaaagg aagaatttcg ggaggccaag	2460
ccccgctccc tccgcttcac gtggagttatg aagaccacga gctccatggaa gcccaacgag	2520
atgatgcggg agatccgcaa ggtgtggac gcgaacagct gccagagcga gctgcattgag	2580
aagtacatgc tgctgtgcac gcacggcactg cggggccacg aggacttcgt gcagtggag	2640
atggaggtgt gcaaactgcc ggggtctct ctcaacgggg ttcgatttaa gcggatatacg	2700
ggcacctcca tggccttcaa aaacattgcc tccaaaatag ccaacgagct gaagctttaa	2760
caggctgcca ggagcggggg cggggggggg cggggccagct ggacgggctg cggcccgctg	2820
cggccccc cctggcgcag actgcagcga tggattggtg tgtctccct gctggcactt	2880
ctccctccc tggcccttct cagttttctc ttacatgttt gtgggggggtg ggagatgtt	2940
ntccagcacc ccacattcac ccctgcccag agattcccccc ttctcctctc ccctactggaa	3000
ggcaaaggaa .ggggaggggtg gatggggggg cagggtctccc cctcggtact gcggttgcac	3060
agagtatttc gcctaaacca agaaattttt tattaccaaaa aag	3103

<210> 4
<211> 2086
<212> DNA
<213> Homo sapiens

<400> 4 agtccaaacat gattcggggc cgcaactcag ccacctctgc ttagtggcagccccacattg	60
gaaaactacccg gtcctcaag accattggca agggtaattt tgccaaagggtg aagttggccc	120
gacacatcct gactggaaa gagtagctg tgaagatcat tgacaagact caactgaact	180
cctccagcct ccagaaacta ttccgcgaag taagaataat gaaggttttg aatcatccca	240
acatagttaa attatggaa gtgattgaga ctgagaaaac gctctacctt gtcattggagt	300
acgcttagtgg cggagaggtt tttgattacc tagtggctca tggcaggatg aaagaaaaag	360
aggctcgagc caaattccgc cagatgtgt ctgctgtgca gtactgtcac cagaagtttta	420
ttgtccatag agacttaaag gcagaaaacc tgctcttggaa tgctgatatg aacatcaaga	480
ttgcagactt tggcttcagc aatgaattca cctttggaa caagctggac accttctgtg	540
gcagtcggcc ttatgctgcc ccagaactct tccagggcaa aaaatatgtat ggaccggagg	600

tggatgtgtg gaggccttagga gttatcctct atacacttgtt cagcgatcc ctgccttttg
atggacagaa cctcaaggag ctgcgggaac gggtaactgag gggaaaatac cgtattccat
tctacatgtc cacggactgt gaaaacctgc ttaagaaaatt ttcattttt aatcccagca
agagaggcac tttagagcaa atcatgaaag atcgatggat gaatgtgggt cacgaagatg
atgaactaaa gccttacgtg gagccactcc ctgactacaa ggacccccgg cgacagagc
tgcgtgtc catggtttat acacgggaag agatccagga ctcgcgtgtg ggccagagat
acaacgaggt gatggccacc tatctgctcc tgggttacaa gagctccgag ctggaaaggcg
acaccatcac cctgaaaccc cggccattcag ctgatctgac caatagcagc gccccatccc
catcccacaa ggtacagcgc agcgtgtcgg ccaatccaa gcagcggcgc ttcagcgcacc
aggctggtcc tgccattccc acctctaatt ttactctaa gaagactcag agtaacaacg
cagaaaataa gcggcctgag gaggaccggg agtcagggcg gaaagccagc agcacagcca
aggtgcctgc cagccccctg cccggtctgg agaggaagaa gaccacccca acccccctcca
cgaacacgcgt cctctccacc agcacaaatc gaagcaggaa ttccccactt ttggagcggg
ccaggcctcgg ccaggcctcc atccagaatg gcaaagacag cacagcccc cagcgtgtcc
ctgttgccctc cccatccgccc cacaacatca gcagcagtgg tggagccca gaccgaacta
acttcccccg gggtgtgtcc agccgaagca cttccatgc tggcagctc cgacaggtgc
gggaccagca gaatttgccc tacggtgtga cccagcctc tccatctggc cacagccagg
gccggcgggg ggcctctggg agcatctca gcaagttcac ctccaagttt gtacgcagga
acctgaatga acctgaaagc aaagaccgag tggagacgct cagacccac gtggtggca
gtggcggcaa cgacaaagaa aaggaagaat ttctggagggc caagccccgc tccctccgt
tcacgtggag tatgaagacc acgagctcca tggagccaa cgagatgtg cggagatcc
gcaagggtgtt ggacgcgaac agctgccaga gcgagctgca tgagaagtac atgctgt
gcatgcacgg cacgccccggc cacgaggact tcgtgcagtg ggagatggag gtgtgcac
tgccgcggct ctctctcaac ggggttcgat ttaagcggat atcgggcacc tccatggcct
tcaaaaaacat tgcctccaaa atagccaaacg agctgaagct ttaaca

<210> 5
<211> 2248
<212> DNA
<213> *Homo sapiens*

```
<400> 5
agtccaaacat gattcggggc cgcaactcag ccacacctgc ttagtggcgag cccccacattg 60
gaaactaccg gctcctcaag accattggca agggtaattt tgccaaagggtg aagttggccc 120
gacacatcct gactggggaaa gaggttagctg tgaagatcat tgacaagact caactgaact· 180
```

cctccagcct ccagaaaacta ttccgcgaag taagaataat gaagggttttgc	240
acatagttaa attatttgaa gtgattgaga ctgagaaaaac gctctacctt gtcatggagt	300
acgcttagtgg cggagaggta tttgattacc tagtggctca tggcaggatg aaagaaaaag	360
aggotcgagc caaattccgc cagatagtgt ctgctgtca gtactgtcac cagaagtttgc	420
ttgtccatag agacttaaag gcagaaaaacc tgctcttggaa tgctgtatgc aacatcaaga	480
ttgcagactt tggcttcagc aatgaattca cctttggaa caagctggac accttctgtgc	540
gcagtcggcc ttatgctgcc ccagaactct tccagggcaa aaaatatgtat ggacccgagg	600
tggatgtgtg gagccttagga gttatcctct atacactggc cagcggatcc ctgcctttgc	660
atggacagaa cctcaaggag ctgcgggaac gggtaactgag gggaaaatac cgtattccat	720
tctacatgtc cacggactgt gaaaacctgc ttaagaaatt tctcattctt aatcccagca	780
agagaggcac ttttagagcaa atcatgaaag atcgatggat gaatgtgggt cacgaagatg	840
atgaactaaa gccttacgtg gagccactcc ctgactacaa ggacccccgg cggacagagc	900
tgatgggtgc catgggttat acacgggaag agatccagga ctcgctggtg ggccagagat	960
acaacgaggt gatggccacc tatctgtcc tggctacaa gagctccgag ctggaaaggcg	1020
acaccatcac cctgaaaccc cggccttcag ctgatctgac caatagcagc gccccatccc	1080
catcccacaa ggtacagcgc agcgtgtcgg ccaatcccaa gcagcggcgc ttcaagcggacc	1140
aggctggtcc tgccattccc acctctaatt cttaactctaa gaagactcag agtaacaacg	1200
cagaaaataa gcggcctgag gaggacccgg agtcagggcg gaaagccagc agcacagcca	1260
aggtgtcgc cagccccctg cccggctctgg agaggaagaa gaccacccca accccctcca	1320
cgaacagcgt cctctccacc agcacaaaatc gaagcaggaa ttccccactt ttggagcggg	1380
ccagcctcgg ccagggctcc atccagaatg gcaaaagacag cctaaccatg ccagggtccc	1440
gggcctccac ggcttctgct totgcccgc tctctgcggc cggccccggc cagcaccaga	1500
aatccatgtc ggccctccgtc caccccaaca aggccctctgg gctgcccccc acggagagta	1560
actgtgaggt gccgcggccc agcacagccc cccagcgtgt ccctgttgc tccccatccg	1620
cccacaacat cagcagcagt ggtggagccc cagaccgaac taacttcccc cgggggtgtgt	1680
ccagccgaag caccttccat gctgggcagc tccgacaggt gcgggaccag cagaatttgc	1740
cctacgggtg gaccccagcc tctccctctg gccacagcca gggccggcgg gggccctctg	1800
ggagcatctt cagcaagttc acctccaagt ttgtacgcag gaacctgaat gaacctgaaa	1860
gcaaagaccc agtggagacg ctcagaccc acgtgggtgg cagtggcggc aacgacaaag	1920
aaaaggaaga atttcgggag gccaagcccc gctccctccg cttcacgtgg agtatgaaga	1980
ccacgagctc catggagccc aacgagatga tgcgggagat cgcgaagggtg ctggacgcga	2040

acagctgcca gagcgagctg catgagaagt acatgctgct gtgcacgcac ggcacgcgg	2100
gccacgagga cttcgtgcag tggagatgg aggtgtcaa actgcgcgg ctctctctca	2160
acggggttcg atttaagcgg atatcggca cttccatggc cttcaaaaac attgcctcca	2220
aaatagccaa cgagctgaag ctttaaca	2248
<210> 6	
<211> 2701	
<212> DNA	
<213> Homo sapiens	
<400> 6	
ggcacgaggg ctgaacgaga gggacacgga gcagcccacc ttggacacc ttgactccaa	60
gccccaggcgt aagtccaaaca tgattcgggg ccgcaactca gccacctctg ctgatgagca	120
gccccacatt ggaaactacc ggctcctcaa gaccattggc aaggtaatt ttgccaaggt	180
gaagttggcc cgacacatcc tgactggaa agaggtatgt gtgaagatca ttgacaagac	240
tcaactgaac tcctccagcc tccagaaact attccgcgaa gtaagaataa tgaaggttt	300
gaatcatccc aacatgtta aattatttga agtgatttag actgagaaaa cgctctacct	360
tgtcatggag tacgcttagtg gcggagaggt atttgattac ctatggctc atggcaggat	420
gaaagaaaaaa gaggctcgag ccaaattccg ccagatagtg tctgctgtgc agtactgtca	480
ccagaagttt attgtccata gagactaaa ggcagaaaaac ctgctcttgg atgctgatat	540
gaacatcaag attgcagact ttggcttcag caatgaattc acctttggga acaagctgga	600
cacccctgtt ggcagtcccc cttatgctgc cccagaactc ttccagggca aaaaatatga	660
tggaccccgag gtggatgtgt ggagcctagg agttatcctc tataactgg tcagcggatc	720
cctgcctttt gatggacaga acctaaggaa gctgcggaa cgggtactga ggggaaaata	780
ccgtattcca ttctacatgt ccacggactg taaaacctg cttaagaaat ttctcatct	840
taatcccagc aagagaggca cttagagca aatcatgaaa gatcgatgga tgaatgtgg	900
tcacgaagat gatgaactaa agccttacgt ggagccactc cctgactaca aggaccccg	960
gcggacagag ctgatggtgtt tacacggaa gagatccagg actcgcttgt	1020
gggcacagaga tacaacgagg tcatggccac ctatctgctc ctggctaca agagctccga	1080
gctggaaggc gacaccatca ccctgaaacc ccggccttca gctgatctga ccaatagcag	1140
cgccccatcc ccatcccaca aggtacagcg cagcgtgtcg gccaatccca agcagcggcg	1200
cttcagcgac caggcagctg gtcctgccc tcccacctct aattcttact ctaagaagac	1260
tcagagtaac aacgcagaaa ataagcggcc tgaggaggac cgggagtcag ggcggaaagc	1320
cagcagcaca gccaagggtgc ctgcccggcc cctgcccgt ctggagagga agaagaccac	1380

cccaaccccc	tccacgaaca	gcgtccttc	caccagcaca	aatcgaaagca	ggaattcccc	1440
acttttggag	cgggccagcc	tcggtcaggc	ctccatccag	aatggcaaag	acagcctaac	1500
catgccaggg	tcccgggct	ccacggcttc	tgatctgccc	gcagtctctg	cggcccccggcc	1560
ccgcccagcac	cagaaatcca	tgtcgccctc	cgtgcaccccc	aacaaggcct	ctgggctgccc	1620
ccccacggag	agtaactgtg	aggtgccgcg	gcccagcaca	gccccccagc	gtgtccctgt	1680
tgccctccca	tccgcccaca	acatcagcag	cagtggtgga	gccccagacc	gaactaactt	1740
ccccccgggt	gtgtccagcc	gaagcacctt	ccatgctggg	cagctccgac	aggtgcggga	1800
ccagcagaat	ttgcccctacg	gtgtgacccc	agcctctccc	tctggccaca	gccaggggccg	1860
gcggggggcc	tctgggagca	tcttcagcaa	gttcacccctcc	aagtttgtac	gcagaaatct	1920
gtctttcagg	tttgcgcagaa	ggaacctgaa	tgaacctgaa	agcaaaagacc	gagtggagac	1980
gctcagaccc	cacgtggtgg	gcagtggcgg	caacgacaaa	gaaaaggaag	aatttcggga	2040
ggccaagccc	cgctccctcc	gcttcacgtg	gagttatgaa	accacgagct	ccatggagcc	2100
caacgagatg	atgcgggaga	tccgcaaggt	gctggacgcg	aacagctgcc	agagcgagct	2160
gcatgagaag	tacatgctgc	tgtgcattgca	cgccacgcgg	ggccacgcgg	acttcgtgca	2220
gtggggagatg	gaggtgtgca	aactgcccgcg	gctctctctc	aacggggttc	gatthaagcg	2280
gatatcgggc	acctccatgg	ccttcaaaaaa	cattgcctcc	aaaatagcca	acgagctgaa	2340
gcttaaacag	gctgccagga	gcgggggcgg	cgggggcggg	ccagctggac	gggctgcgg	2400
ccgctgcgcc	gccccacctg	ggcgagactg	cagcgatgga	ttggtgtgtc	tccctgtcg	2460
gcacttctcc	cctccctggc	ccttctcagt	tttctcttac	atgtttgtgg	gggggtgggag	2520
attgttctcc	agccccccac	attcaccct	gcccagagat	tcccccttat	cctctccccct	2580
actggaggca	aaggaaggggg	agggtggatg	ggggggcagg	gctcccccctc	ggtactgcgg	2640
ttgcacagag	tatttcgctt	aaacccaagaa	attttttatt	accaaaaaaaaa	aaaaaaaaaaa	2700
a						2701

<210> 7
 <211> 2112
 <212> DNA
 <213> Homo sapiens

<400> 7	cccagcagta	agtccaaacat	gattcggggc	cgcaactcag	ccacctctgc	tgatgagcag	60
	ccccacattg	gaaactaccg	gctcctcaag	accattggca	agggttaattt	tgccaaagggt	120
	aagttggccc	gacacatcct	gactggggaaa	gaggttagctg	tgaagatcat	tgacaagact	180
	caactgaact	cctccagcct	ccagaaacta	ttccgcgaag	taagaataat	gaaggtttg	240
	aatcatccca	acatagttaa	attatggaa	gtgattgaga	ctgagaaaac	gctctacctt	300

gtcatggagt acgctagtgg cggagaggta tttgattacc tagtggctca tggcaggatg	360
aaagaaaaag aggctcgagc caaatccgc cagatagtgt ctgctgtgca gtactgtcac	420
cagaagttta ttgtccatag agacttaaag gcagaaaacc tgcttggaa tgctgatatg	480
aacatcaaga ttgcagactt tggcttcagc aatgaattca cctttggaa caagctggac	540
accttctgtg gcagtcggcc ttatgctgcc ccagaactct tccagggcaa aaaatatgtat	600
ggacccgagg tggatgtgtg gagccttagga gttatcctct atacactggc cagcggatcc	660
ctgccttttgc atggacagaa cctcaaggag ctgcgggaaac gggtaactgag gggaaaatac	720
cgtattccat tctacatgtc cacggactgt gaaaacctgc ttaagaaatt ttcattttt	780
aatcccagca agagaggcac ttttagagcaa atcatgaaag atcgatggat gaatgtgggt	840
cacgaagatg atgaactaaa gccttacgtg gagccactcc ctgactacaa ggaccccccgg	900
cggacagagc tgatgggtgc catgggttat acacgggaaag agatccagga ctcgctggtg	960
ggccagagat acaacgaggt gatggccacc tatctgctcc tgggctacaa gagctccgag	1020
ctggaaggcg acaccatcac cctgaaaccc cggccttcag ctgatctgac caatagcagc	1080
ccccatcccc catccccacaa ggtacagcgc agcgtgtcgg ccaatcccaa gcagcggcgc	1140
ttcagcgcacc aggctggtcc tgccattccc acctctaatt cttactctaa gaagactcag	1200
agtaacaacg cagaaaataa gcggcctgag gaggaccggg agtcagggcg gaaagccagc	1260
agcacagcca aggtgcctgc cagccccctg cccggctcgg agaggaagaa gaccacccca	1320
accccccctcca cgaacagcgt cctctccacc agcacaaatc gaagcaggaa ttccccactt	1380
ttggagcggg ccagcctcgg ccaggcctcc atccagaatg gcaaagacag cacagcccc	1440
cagcgtgtcc ctgttgccctc cccatccgccc cacaacatca gcagcagtgg tggagcccc	1500
gaccgaacta acttcccccg ggggtgttcc agccgaagca cttccatgc tggcagctc	1560
cgacaggtgc gggaccagca gaatttgcctc tacgggtgtga ccccagccctc tccctctggc	1620
cacagccagg gccggcgcccc ggcctctggg agcatcttca gcaagttcac ctccaaatgttt	1680
gtacgcagga acctgaatga acctgaaagc aaagaccgag tggagacgct cagacccatcac	1740
gtgggtggca gtggcggcaa cgacaaagaa aaggaagaat ttcgggaggc caagccccgc	1800
tccctccgct tcacgtggag tatgaagacc acgagctcca tggagcccaa cgagatgtat	1860
cgggagatcc gcaaggtgct ggacgcgaac agctgccaga gcgagctgca tgagaagtac	1920
atgctgtgt gcatgcacgg cacggccggcc cacgaggact tcgtgcagtgc ggagatggag	1980
gtgtgcaaac tgccgcggct ctctctcaac ggggttcgat ttaagcggat atcgggacc	2040
tccatggcct tcaaaaacat tgccctccaaa atagccaaacg agctgaagct ttaacaggct	2100
gccaggagcg gg	2112

<210> 8
<211> 2965
<212> DNA
<213> Homo sapiens

<400> 8						
cgggcaaccg	cctcgccgaa	agccctccct	cgttactgtc	cgcataacccc	ggcgccgccc	60
ccgcggaaag	cggctcccc	tcctcttaact	ccgcgtactc	ttccctcttt	ccccccgg	120
ggcacgcttgc	ttgcaccgtc	ccgcggcctg	cgggagccgc	tcgccccgga	cgttagctcg	180
cgtacgaccc	atttcctgtc	gcccccggga	gcccgcacca	cagccggcc	ggtctagacc	240
ccggcagacc	ccgctggccg	cacaaaatgt	cggccggac	gccattgccc	acggtaac	300
agcgggacac	gttaaatcat	acgactgtgg	atggatatac	tgaaccacac	atccagccta	360
ccaagtcgag	tagcagacag	aacatcccc	ggtgtagaaa	ctccattacg	ttagcaacag	420
atgaacagcc	tcacattgga	aattaccgtt	tacaaaaaac	aatagggaag	ggaaattttg	480
ccaaagtcaa	attggcaaga	cacgttctaa	ctggtagaga	ggttgctgt	aaaataatag	540
acaaaactca	gctaaatcct	accagtctac	aaaagttatt	tcgagaagta	cgaataatga	600
agatactgaa	tcatccta	atagtaaaat	tgtttgaagt	tattgaaaca	gagaagactc	660
tctatttatgt	catggaatac	gcgagtgggg	gtgaagtatt	tgattactta	gttgcccatg	720
gaagaatgaa	agagaaagag	gcccgtaaa	aatttaggca	gattgtatct	gctgtacagt	780
attgtcatca	aaagtacatt	gttcaccgtg	atcttaaggc	tgaaaacctt	ctcattgtatg	840
gtgatatgaa	tataaaaatt	gctgactttg	gttttagtaa	tgaatttaca	gttggaaaca	900
aattggacac	attttgtgga	agccaccct	atgctgtcc	cgagctttc	caaggaaaga	960
agtatgatgg	gcctgaagtg	gatgtgtgga	gtctggcgt	cattctctat	acattagtca	1020
gtggctcatt	gcctttcgat	ggccagaatt	taaaggaact	gcgagagcga	gttttacgag	1080
ggaagtaccg	tattcccttc	tatatgtcca	cagactgtga	aatcttctg	aagaaattat	1140
tagtcctgaa	tccaataaaag	agaggcagct	tggacaat	aatgaaagat	cgatggatga	1200
atgttggtca	tgaagagggaa	gaactaaagc	catatactga	gcctgatccg	gatttcaatg	1260
acacaaaaag	aatagacatt	atggtcacca	tggccttgc	acgagatgaa	ataaaatgtat	1320
ccttaataaa	tcagaagtat	gatgaagtta	tggctactta	tattcttcta	ggtagaaaac	1380
cacctgaatt	tgaaggtgg	gaatcgat	ccagtggaaa	cttgtgtcag	aggtccggc	1440
ccagtagtga	cttaaacaac	agcactcttc	agtccccgtc	tcacctgaag	gtccagagaa	1500
gtatctcagc	aaatcagaag	cagcggcg	tcagtgtatca	tgctggtcca	tccattcctc	1560
ctgctgtatc	atataccaaa	agacctcagg	ctaacagtgt	ggaaagtgaa	cagaaagagg	1620

agtgggacaa agatgtggct cgaaaacttg gcagcacaac agttggatca aaaagcgaga	1680
tgactgcaag ccctcttgc gggccagaga ggaaaaaaatc ttcaactatt ccaagtaaca	1740
atgtgtattc tggaggtagc atggcaagaa ggaatacata tgtctgtgaa aggaccacag	1800
atcgatacgt agcattgcag aatggaaaag acagcagcct tacggagatg tctgtgagta	1860
gcatatcttc tgcaggctct tctgtggct ctgctgtccc ctcagcacga ccccgccacc	1920
agaagtccat gtccacttct ggtcatccta ttaaaagtac actgccaacc attaaagacg	1980
gctctgaagc ttaccggct ggtacaaccc agagagtgcc tgctgttcc ccattctgctc	2040
acagtattag tactgcgact ccagaccgga cccgtttcc ccgagggagc tcaagccgaa	2100
gcactttcca tggtaaacag ctccgggagc gacgcagcgt tgcttataat gggccacctg	2160
cttcaccatc ccatgaaacg ggtgcatttgc acatgccag aagggaaacg tcaactggta	2220
taataagcaa aatcacatcc aaatttgcgc gcagggatcc aagtgaaggc gaagccagtg	2280
gcagaaccga cacctcaaga agtacatcg gggAACAAA agaaagagac aaggaagagg	2340
gtaaagattc taagccgcgt tctttgcggc tcacatggag tatgaagacc actagttcaa	2400
tggaccctaa tgacatgatg agagaaaatcc gaaaagtgtt agatgcaaat aactgtgatt	2460
atgagcaaaa agagagattt ttgccttct gtgtccatgg agacgctaga caggatagcc	2520
tcgtgcagtg ggagatggaa gtctgcaagt tgccacgact gtcacttaat ggggttcgct	2580
tcaagcgaat atctggaca tctattgcct ttaagaacat tgcatcaaaa atagcaaatg	2640
agcttaagct gtaaagaagt ccaaattac aggttcaggg aagatacata catatatgag	2700
gtacagtttt tgaatgtact ggtaatgcct aatgtggct gcctgtgaat ctccccatgt	2760
agaatttgccttaatgcaaa taaggttata catagttatg aactgtaaaa ttAAAGTCAG	2820
tatgaactat aataaatatc ttagctaa aaagtaggtt cacatgtaca ggtaagtata	2880
tttgttattt ctgttcattt tctgttcata gagttgtata ataaaacatg attgcttaaa	2940
aacttgaaaa aaaaaaaaaaaa aaaaaa	2965

<210> 9
 <211> 3210
 <212> DNA
 <213> Homo sapiens

<400> 9	
ggcgccggcgg cggcggtggc tgtgaccgcg cggaccgagc cgagacattc gcgccgggggg	60
atcgccggccgccc gccgcccgtg ggccccgggc gcggtggatgc ggctgggtcg ggccggcccg	120
tacacctgag gcgagaaacg gggcgccggcg cgggtgacgc tgtcagggcc gcggttcctg	180
acgcccaggc gctcgccagg acgagccagg cagtgttttgc aggcaccggc ttcacctca	240
cccatggtcc ggagagccta gcggggctcg ccaccgcctc ccggctcccc ttccacgcct	300

catcctgccca	gcctcgccgc	cccggccagcg	ccggggcaacc	gcctcgccccg	aagccctccc	360
tatgttactgt	ccgcataacc	cggcgccggcc	gccgcgggaa	gcggctcccc	cttcctttcc	420
tccgcgttct	tttccttttt	tcccccggcg	ggggccgattt	ttgcacccgcc	ccgcggccctg	480
ccggagccgc	tcgccccggc	tttgtgtcg	cgtccgcacc	cctttctgt	cgccccccgg	540
ggcccgccacc	acagccccggc	cggcgagacc	ccggccagac	cccgctgccc	gcacaaaaatg	600
tccggcccgga	cgccattgcc	gacggtaaac	gagcgggaca	cggaaaatca	tacatctgtg	660
gatggatata	ctgaaccaca	catccagcct	accaagtgcg	gtagcagaca	gaacatcccc	720
cggtagaa	actccattac	gtcagcaaca	gatgaacagc	ctcacattgg	aaattaccgt	780
ttacaaaaaa	caataggaa	gggaaatttt	gccaaagtca	aattggcaag	acacgttcta	840
actggtagag	aggttgtgt	gaaaataata	gacaaaactc	agctaaatcc	taccagtcta	900
caaaaagttat	ttcgagaagt	acgaataatg	aagataactga	atcatcctaa	tataggtgaa	960
gtatttgatt	acttagttgc	ccatggaaaga	atgaaagaga	aagaggcccc	tgcaaaattt	1020
aggcagattg	tatctgctgt	acagtattgt	catcaaaagt	acattgttca	ccgtgatctt	1080
aagctgaaaa	ccttctcctt	gatggtgata	tgaatattaa	aattgctgac	tttggtttta	1140
gtaatgaatt	tacagttggg	aacaaattgg	acacattttg	tggaaagccca	ccctatgctg	1200
ctcccgagct	tttccaagga	aagaagtatg	atggccctga	agtggatgtg	tggagtgctgg	1260
gcgtcattct	ctatacatta	gtcagtggtc	ccttgccttt	cgatggccag	aatttaaagg	1320
aactgcgaga	gcgagtttta	cgagggaaagt	accgtattcc	cttcttatatg	tccacagact	1380
gtgaaaatct	tctgaagaaa	ttatttagtcc	tgaatccaat	aaagagagggc	agcttggAAC	1440
aaataatgaa	agatcgatgg	atgaatgttg	gtcatgaaga	ggaagaacta	aagccatata	1500
ctgagcctga	tccggatttc	aatgacacaa	aaagaataga	cattatggtc	accatgggct	1560
ttgcacgaga	tcaaataaaat	gatgccttaa	taaatcagaa	gtatgatgaa	gttatggcta	1620
cattatattct	tcttaggtaga	aaaccacctg	aatttgaagg	tggtaatcg	ttatccagtg	1680
gaaacttgtg	tcaagaggccc	cggcccagta	gtgacttaaa	caacagcact	cttcagtc	1740
ctgctcacct	gaaggtccag	agaagtatct	cagcaaata	gaagcagcgg	cgtttcagtg	1800
atcatgctgg	tccatccatt	cctcctgctg	tatcatatac	caaaaagacct	caggctaaaca	1860
gtgtggaaag	tgaacagaaa	gaggagtggg	acaaagatgt	ggctcgaaaa	cttggcagca	1920
caacagttgg	atcaaaaagc	gagatgactg	caagccctct	tgttagggcca	gagaggaaaa	1980
aatcttcaac	tattccaagt	aacaatgtgt	attctggagg	tagcatggca	agaaggaata	2040
catatgtctg	tgaaaggacc	acagatcgat	acgttagcatt	gcagaaatgga	aaagacagca	2100
gccttacgga	gatgtctgt	agtagcatat	cttctgcagg	ctcttctgtg	gcctctgtg	2160

tccccctcagc acgaccccgcc	caccagaagt ccatgtccac ttctggtcat	cctattaaag	2220
tcacactgcc aaccattaaa gacggctctg	aagcttacccg gcctggtaca	acccagagag	2280
tgcctgctgc ttccccatct	gctcacagta ttagtactgc	gactccagac cggacccgtt	2340
ttccccgagg gagctcaagc	cgaagcactt tccatggtga	acagctccgg gagcgacgca	2400
gggttgctta taatgggcca	cctgcttcac catcccatga	aacgggtgca tttgcacatg	2460
ccagaagggg aacgtcaact	ggtataataa gcaaaatcad	atccaaattt gttcgacagg	2520
atccaaatgtga aggcgaaagcc	agtggcagaa ccgacaccc	aagaagtaca tcaggggaac	2580
caaaagaaaag agacaaggaa	gagggtaaag attctaagcc	gcgttctttg cggttcacat	2640
ggagtatgaa gaccactagt	tcaatggacc ctaatgacat	gatgagagaa atccgaaaag	2700
tgttagatgc aaataactgt	gattatgagc aaaaagagag	atttttgctt ttctgtgtcc	2760
atggagacgc tagacaggat	agcctcggtgc agtggagat	ggaagtctgc aagttgccac	2820
gactgtcact taatggggtt	cgcttcaagc gaatatctgg	gacatctatt gcctttaaga	2880
acattgcac	aaaaatagca aatgagctt	agctgtaaag aagtccaaat ttacaggttc	2940
agggaaagata catacatata	tgaggtacag ttttgaatg	tactggtaat gcctaatttg	3000
gtctgcctgt	aatctcccc atgtagaatt	tgcccttaat gcaataaggt tatacatagt	3060
tatgaactgt	aaaattaaag tcagtatgaa	ctataataaa tatctgtac ttaaaaagta	3120
ggttcacatg tacaggtaa	tatattgtgt	atttctgttc attttctgtt catagagtt	3180
tataataaaa	catgattgct	taaaaaacttg	3210

<210> 10
<211> 2505
<212> DNA
<213> Homo sapiens

<400> 10	60					
gctggccgca caaaatgtcg	gccggacgc cattgccac	ggtaacgag cggacacgg	60			
aaaatcatac atctgtggat	ggatatactg	aaccacacat ccagctacc	aagtcgagta	120		
gcagacagaa catccccgg	tgtagaaaact	ccattacgtc	agcaacagat	gaacagccctc	180	
acattggaaa ttaccgtta	caaaaaacaa	taggaaaggg	aaattttgcc	aaagtcaaat	240	
tggcaagaca cgttctaact	ggttagagagg	ttgctgtgaa	aataatagac	aaaactcagc	300	
taaatcctac cagtctacaa	aagttatitc	gagaagtacg	aataatgaag	atactgaatc	360	
atcctaataat	agtaaaaattt	tttgaagtta	ttgaaacaga	gaagactctc	tatattgtca	420
tggaatacgc	gagtgggggt	gaagtatttgc	attacttagt	tgcccatgga	agaatgaaag	480
agaaaagaggc	ccgtgcaaaa	tttaggcaga	ttgtatctgc	tgtacagtat	tgtcatcaa	540

agtacattgt tcaccgtgat cttaaggctg aaaaccttct ccttgatggt gatatgaata	600
ttaaaattgc tgacttttgt ttttagtaatg aatttacagt tggaacaaa ttggacacat	660
tttgtgaaag cccaccctat gctgctcccg agctttcca aggaaagaag tatgatggc	720
ctgaagtgga tgtgtggagt ctggcggtca ttctctatac attagtcagt ggctccttgc	780
cttcgatgg ccagaattta aaggaactgc gagagcgagt ttacgaggg aagtaccgt	840
ttcccttcta tatgtccaca gactgtgaaa atcttctgaa gaaattatta gtctgaatc	900
caataaaagag aggcagcttgaacaaataa tgaaagatcg atggatgaat gttggtcatg	960
aagaggaaga actaaagcca tatactgagc ctgatccgga ttcaatgac aaaaaaagaa	1020
tagacattat ggtcaccatg ggcttgcac gagatgaaat aaatgatgcc ttaataaaatc	1080
aagaagtatga tgaagttatg gctacttata ttcttctagg tagaaaaacca cctgaatttg	1140
aaggtggtga atcggttatcc agtggaaact tgtgtcagag gtcccggccc agtagtgact	1200
taaacaacag cactttcag tccccgtc acctgaaggt ccagagaagt atctcagcaa	1260
atcagaagca gcggcggttc agtgatcatg ctggccatc cattccctt gctgtatcat	1320
atacaaaaag acctcaggct aacagtgtgg aaagtgaaca gaaagaggag tgggacaaag	1380
atgtggctcg aaaacttggc agcacaacag ttggatcaaa aagcgagatg actgcaagcc	1440
ctcttctagg gccagagagg aaaaaatctt caactattcc aagtaacaat gtgtattctg	1500
gaggttagcat ggcaagaagg aatacatatg tctgtgaaag gaccacagat cgatacgtag	1560
cattgcagaa tggaaaagac agcagccctt cggagatgtc tgtgagtagc atatcttctg	1620
caggcttttc tgtggcctct gctgtcccc cagcacgacc ccgcccaccag aagtccatgt	1680
ccacttctgg tcattcttatt aaagtacacac tgccaaaccat taaagacggc tctgaagctt	1740
acggccctgg tacaacccag agagtgcctg ctgttcccc atctgctcac agtattagta	1800
ctgcgactcc agaccggacc cgttttcccc gagggagctc aagccgaagc actttccatg	1860
gtgaacagct ccgggagcga cgcagcggtt cttataatgg gccacctgct tcaccatccc	1920
atgaaaacggg tgcatttgca catgccagaa gggaaacgtc aactggtata ataagcaaaa	1980
tcacatccaa atttgttcgc agggatccaa gtgaaggcga agccagtggc agaaccgaca	2040
cctcaagaag tacatcaggg gaaccaaaag aaagagacaa ggaagagggt aaagattcta	2100
agccgcgttc tttgcgggttc acatggagta tgaagaccac tagttcaatg gaccctaatg	2160
acatgatgag agaaatccga aaagtgttag atgcaaataa ctgtgattat gagcaaaaag	2220
agagattttt gctttctgt gtccatggag acgctagaca ggatagccctc gtgcagtgg	2280
agatggaagt ctgcaagttg ccacgactgt cacttaatgg gttcgcttc aagcgaatat	2340
ctgggacatc tattgcctt aagaacatttgc catcaaaaat agcaaatgag cttaagctgt	2400

aaagaagtcc aaatttacag gttcaggaa gatacataca tatacgatgt acagtttttg	2460
aatgtactgg taatgectaa tgtggctgc ctgtgaatct cccca	2505

<210> 11
<211> 4638
<212> DNA
<213> Homo sapiens

<400> 11	
ggcgccggcgg cggcggtggc tgtgaccgcg cggaccgagc cgagacattc gcgccgggggg	60
atcggggcgcc gcccggcgtg ggccccgggc gcgtggatgc ggctgggtcg ggcggcgcccg	120
tacacctgag gcgaggagaacg gggcgccggc cgggtgacgc tgtcagggcc gcggttcctg	180
acgcccaggc gtcgcccagg acgagccagg cagtgatttgc aggccacggc ttcaccttca	240
cccatggtcc ggagagcccta gcggggctcg ccaccgcctc ccggctcccc ttccacgcct	300
catcctgcca gcctcgccgc cccggcagcg cggggcaacc gcctcgcccg aagccctccc	360
tcgttactgt ccgcatacccg cggcgccggc gccgcgggaa gcggctcccc ctccctttcc	420
tccgcgtcct cttcccttcc tccccggcgg gggccgcttg ttgcacccgc ccggggactg	480
cgggagccgc tcgccccggc cttgtgctcg cgtccgcacc cctttctgt cgccccccgg	540
ggcccgccacc acagcccgcc cggcgagacc cggccagac cccgctgccc gcacaaaatg	600
tcggcccgga cgcattgcc gacggtaac gagcgggaca cggaaaatca tacatctgt	660
gatggatata ctgaaccaca catccagcct accaagtcga gtagcagaca gaacatcccc	720
cggtagaa actccattac gtcagcaaca gatgaacacgc ctcacattgg aaattaccgt	780
ttacaaaaaa caataggaa gggaaatttt gccaaagtca aattggcaag acacgttcta	840
actggtagag aggttgctgt gaaaataata gacaaaactc agctaaatcc taccagtcta	900
caaaagttat ttcgagaagt acgaataatg aagataactga atcatcctaa tataggtaa	960
gtatggatt acttagttgc ccatggagaatg atgaaagaga aagaggcccg tgcaaaaattt	1020
aggcagattt tatctgctgt acagtattgt catcaaaagt acattgttca cctgtatctt	1080
aaggctgaaa accttctcct ttagggtgat atgaatatta aaattgttca ctttgggttt	1140
agtaatgaat ttacagttgg gaacaaattt gacacatttt gtggaaagccc accctatgt	1200
gctcccgagc ttttccaagg aaagaagtat gatgggcctg aagtggatgt gtggagtcgt	1260
ggcgcatcc tctatacatt agtcagtggc tccttgcctt tcgtatggcca gaatttaaag	1320
gaactgcgag agcgagttt acgaggaaag taccgtattt ctttctatat gtccacagac	1380
tgtgaaaatc ttctgaagaa attattagtc ctgaatccaa taaagagagg cagcttggaa	1440
caaataatga aagatcgatg gatgaatgtt ggtcatgaag aggaagaact aaagccatat	1500
actgagcctg atccggattt caatgacaca aaaagaatag acattatgtt caccatggc	1560

tttgcacgag atgaaataaa ttagtcctta ataaatcaga agtatgtga agttatggct 1620
acttatattc ttcttaggtag aaaaccacct gaatttgaag gtggtaatc gttatccagt 1680
ggaaaacttgt gtcagaggc ccggcccagt agtgacttaa acaacagcac tccttcgtcc 1740
cctgctcacc tgaaggtcca gagaagtatc tcagcaaatac agaagcagcg gcgtttcagt 1800
gatcatgtg gtccatccat tcctcctgt gtatcatata ccaaaagacc tcaggctaac 1860
agtgtggaaa gtgaacagaa agaggagtgg gacaaagatg tggctcgaaa acttggcagc 1920
acaacagttg gatcaaaaag cgagatgact gcaaggccctc ttgttagggcc agagagggaaa 1980
aaatcttcaa ctattccaag taacaatgtg tattctggag gtagcatggc aagaaggaat 2040
acatatgtct gtgaaaggac cacagatcga tacgttagcat tgcagaatgg aaaagacagc 2100
agccttacgg agatgtctgt gagtagcata tccttcgtcag gctttctgt ggccctctgt 2160
gtccccctcag cacgaccccg ccaccagaag tccatgtcca cttctggtca tccttattaaa 2220
gtcacactgc caaccattaa agacggctct gaagcttacc ggccctggtac aacccagaga 2280
gtgcctgctg cttccccatc tgctcacagt attagtaactg cgactccaga cccggacccgt 2340
tttccccgag ggagctcaag ccgaaggact ttccatggtg aacagctccg ggagcgacgc 2400
agcggttgc ttataatggcc acctgcctca ccatcccatg aaacgggtgc atttgcacat 2460
gccagaaggg gaacgtcaac tggtataata agcaaaatca catccaaatt tgttcgcaga 2520
agtacatcag gggAACCAAA agaaagagac aaggaagagg gtAAAGATTc taagccgcgt 2580
tctttgcggc tcacatggag tatgaagacc actagttcaa tggaccctaa tgacatgatg 2640
agagaaatcc gaaaagtgtt agatgcaaat aactgtgatt atgagcaaaa agagagattt 2700
ttgcctttct gtgtccatgg agacgctaga caggatagcc tcgtgcagtg ggagatggaa 2760
gtctgcaagt tgccacgact gtcacttaat ggggttcgt tcaagcgaat atctgggaca 2820
tctattgcct ttaagaacat tgcatcaaaa atagcaaatac agcttatgct gtAAAGAAGT 2880
ccaaatttac aggttcaggg aagatacata catatatgag gtacagttt tgaatgtact 2940
ggtaatgcct aatgtggct gcctgtgaat ctccccatgt agaatttgc cttaatgc当地 3000
taaggttata catagttatg aactgtaaaa ttaaagtcaag tatgaactat aataaatatc 3060
tgttagctaa aaagttagtt cacatgtaca ggtaagtata ttgtgtatTT ctgttcattt 3120
tctgttcata gagttgtata ataaaacatg attgcttaaa aacttgtata gttgtctaga 3180
tttctgcacc tgaatgtatg tttgatgtct tgatttgaaa atgttttcc ctgttattta 3240
cattctggtg ggtttttaaa attcttacct ccatcatgca attttggaaa ttgtgtccag 3300
aattaaaagt gcatagaaaat agcctttaca attgttagcat ggaccctttaa aaattgtttt 3360
aaaatcttat ttaaattttaa accagaagct gaaaaataga tcagctttat tatacacaaa 3420

attattactg	cttatcttt	ctctttcct	tgttatcccg	caaggtttag	ttgagaagat	3480
acaaaatgtt	tacagtgtt	gcacttagag	ttttaaatt	caagtacatg	aaattcagta	3540
atagcattgc	cttgagctaa	ctaggaagta	ccggaaaaaa	agttaaatct	acatcaagtt	3600
tctttgaac	tttgaagtgt	tttctgaccc	actgctaact	gtagcaacaa	aatttaaaag	3660
aaaaaaaaaca	tactttatct	ggctattata	acataaactg	tcacgttagt	ttgctgcctt	3720
cagaataccg	caatttaatt	gcggaaatat	aataatattg	ggactgtttc	acagcacaaa	3780
ctcatctta	cagtgtt	caatgcatca	gttaagaaat	aatgccacct	caggaattaa	3840
ctggcattgg	gaacatttgc	ctcattctcc	tgctatcctc	ttcattcacc	cctgccactg	3900
taatatctat	aagtacttaa	gagacttgt	agcaaaacat	actattata	acagtatatg	3960
attgatttat	gcttatgtgg	ttgttcagtt	tgttccatg	taactcgttt	gttttaata	4020
ttttgccaga	tttcttgtat	ttattccaca	tcattatgcc	tataatgtgc	cgctttgtga	4080
ttggcattt	gcctactttt	ctttcataat	tagtgatata	tgcgatgtaa	aaccactagt	4140
aaaggta	tttaataactt	gttattttat	actgaattag	ccttggaggt	tgactgtgca	4200
atgttattta	ctgtt	taactgtaata	ccaacatatg	ggccccatct	gcacactcct	4260
aaaaaacaga	aagtgtattc	aaattttatc	agtttaaaga	aaataaagct	gtgataaaata	4320
ctgtaattcc	aacctacatt	agaaggtcta	agtgttaggt	atgtgccatt	ccataatggc	4380
ttccagacta	gggtgaattt	tatgttctgt	actgtactgt	gatgtagctt	tcttctgtaa	4440
cagttatgtt	ttaaaattaa	gtgagttttt	tttttgcctt	agcaaagggt	ggtgtttgaa	4500
aaaaaaaaatg	tgtagcccct	tttaaccta	gtgttcattc	aaaaaaaaat	tgatgcaaata	4560
ctttattcac	tttca	ctgttgcgtt	gcacactgaa	atttacttg	aacagttctc	4620
acttgtcttt	tgctcttt					4638

<210> 12
 <211> 2720
 <212> DNA
 <213> Homo sapiens

<400> 12	tcatggaata	cgcgagtggg	ggtgaagtat	ttgattactt	agttgcccatt	ggaagaatga	60
	aagagaaaaga	ggcccggtgca	aaatttaggc	agattgtatc	tgctgtacag	tattgtcattc	120
	aaaagtacat	tgttcaccgt	gatcttaagg	ctgaaaacct	tctcattgtat	ggtgatatga	180
	atattaaaat	tgctgacttt	ggttttagta	atgaatttac	agttggaaac	aaattggaca	240
	cattttgtgg	aagcccaccc	tatgttgctc	ccgagctttt	ccaaggaaag	aagtatgtg	300
	gtcctgaagt	ggatgtgtgg	agtctggcgc	tcattctcta	tacatttagtc	agtggctct	360

tgccttcga tggccagaat ttcaaaggaac tgcgagagcg agttttacga g	420
gtattccctt ctatatgtcc acagactgtg aaaatcttct gaagaaaatta t	480
atccaataaa gagaggcagc ttggaacaaa taatgaaaga tcgatggatg a	540
atgaagagga agaactaaag ccatatactg agcctgatcc ggatttcaat g	600
aatagacat tatggtcacc atgggcttg cacgagatga aataaatgat g	660
atcagaagta tcatgtttttt atggttcttctt aggttagaaaa ccacactgaat	720
ttgaaggtgg tgaatcgta tccagtggaa acttgtgtca gaggtcccgg c	780
ctttaaacaa cagcactctt cagtcctcg ctcacctgaa ggtccagaga agt	840
caaatcagaa gcagcggcgt ttcagtgtatc atgctggtcc atccattcct c	900
catataccaa aagacctcag gctaaccatg tggaaagtga acagaaaagag g	960
aagatgtggc tcgaaaactt ggcagcacaa cagttggatc aaaaagcgag at	1020
gcccttctgtt agggccagag agaaaaaaat cttcaactat tccaagtaac a	1080
atgtgttcatg catggcaaga aggaatacat atgtctgtga aaggaccaca g	1140
tagcattgca gaatggaaaa aacagcagcc ttacggagat gtctgtgagt ag	1200
ctgcaggctc ttctgtggcc tctgctgccc cctcagcacg accccgccac c	1260
tgtccacttc tggtcatcct attaaagtca cactgccaac cattaaagac gg	1320
cttaccggcc tggtacaacc cagagagtgc ctgctgttc cccatctgct c	1380
gtactgcgac tccagaccgg acccgtttc cccgagggag ctcaagccga ag	1440
atggtaaca gctccggag cgacgcagcg ttgcttataa tggccacct g	1500
ccatgaaac gggtgcattt gcacatgcca gaagggaaac gtcaactggg at	1560
aaatcacatc caaatttgtt cgcagaagta catcagggaa accaaaagaa ag	1620
aagaggtaa agattctaag ccgcgttctt tgcggttcac atggagtatg a	1680
ttcaatgga ccctaattgac atgatgagag aaatccaaa agtgttagat g	1740
gtgattatga gcaaaaagag agattttgc ttttctgtgtt ccatggagac g	1800
atagcctcgt gcagtggag atgaaagtct gcaagttgca cgactgtcac t	1860
tcgcttcaag cgaatatctg ggacatctat tgccttaag aacattgcat ca	1920
aaatgagctt aagctgtaaa gaagtccaaa tttacaggtt cagggaaagat ac	1980
atgaggtaca gttttgaat gtactggtaa tgcctaattgt ggtctgcctg t	2040
catgtagaat ttgccttaa tgcaataagg ttatacatag ttatgaactg taaa	2100
gtcagtatga actataataa atatctgttag cttaaaaagt aggttcacat gt	2160
gtatattgtt tatttcgtt cattttctgt tcatagagtt gtataataaa acat	2220

ttaaaaactt gtatagttgt ctagatttct gcacctgaat gtagtttgat	2280
tgaaaatgtt ctccctgtt attacattc cggtgggtt taaaattct tacccatc	2340
atgcaatttt gaaaattgtg tccagaatta aaagtgcata gaaatagcct ttacaattgt	2400
agcatggacc tttaaaaatt gttttaaaat cttatttaaa. tttaaaccag aagctgaaaa	2460
atagatcagc tttattatac aaaaaattat tactgcttat ctttgctctt ttcccttgtt	2520
tcccgcaagg tttagtttag aagataaaaa atgtttacag tggtggcact tagagtttt	2580
aaattcaagt acatgaaatt cagtaatagc attgccttga gctaactagg aagtaccggg	2640
aaaaaaagtta aatctacatc aagtttctt tgaactttga agtgtttctt gaccactgc	2700
taactgttagc aacaaaattt	2720

<210> 13
<211> 2698
<212> DNA
<213> Homo sapiens

<400> 13	
gagctgaaat tcgcgggtcg acggggaggga gtggagaagg aggtgagggg gcccaggatc	60
gcggggcgcc ctgaggcaag gggacgcccgg tgggtcgaag cgcaagccgc cgcccgccagg	120
ctcggtcccg ccactgcccgc ctcgggttc tccctgcctc gggcgccgag gcagggagag	180
aatgagcccc gggacccgcc gggggacggc cggggccagg cccgggatct agaacggccg	240
taggggaag ggagccgccc tccccacggc gcctttcgg aactgcgtg gactcgagga	300
cgctggtcgc cggcctccta gggctgtct gtttttttt gaccctcgca ttgtgcagaa	360
ttaaaagtgcgtca gtaaaatgtc cactaggacc ccattgcca cggtaatga acgagacact	420
aaaaaccaca cgtcacatgg agatggcgt caagaagtta cctctcgta cagccgctca	480
ggagctcggt gtagaaactc tatagcctcc tgtgcagatg aacaacctca catcgaaac	540
tacagactgt tgaaaacaat cggcaaggaa aattttgc当地 aagtaaaatt ggcaagacat	600
atccttacag gcagagaggt tgcaataaaaa ataattgaca aaactcagtt gaatccaaca	660
agtctacaaa agcttttag agaagtaaga ataatgaaga ttttaatca tcccaatata	720
gtgaagttat tcgaagtcat tgaaactgaa aaaacactct acctaattcat ggaatatgca	780
agtggaggtg aagtatttga ctatgggtt gcacatggca ggatgaagga aaaagaagca	840
agatctaaat ttagacagat tgtgtctgca gttcaataact gccatcagaa acggatcgta	900
catcgagacc tcaaggctga aaatctattt gtagatgccg atatgaacat taaaatagca	960
gatttcgggtt ttagcaatga atttactgtt ggcggtaaac tcgacacgtt ttgtggcagt	1020
cctccatacg cagcacctga gctctccag ggcaagaaa atgacgggcc agaagtggat	1080
gtgtggagtc tgggggtcat tttatacaca ctagtcagtg gtcacttcc ctttgatggg	1140

caaaaacctaa aggaactgag agagagagta ttaagaggga aatacagaat tcccttctac 1200
 atgtctacag actgtgaaaa ccttctcaaa cgtttcctgg tgctaaatcc aattaaacgc 1260
 ggcactctag agcaaatcat gaaggacagg tggatcaatg cagggcatga agaagatgaa 1320
 ctcaaaccat ttgttgaacc agagctagac atctcagacc aaaaaagaat agatattatg 1380
 gtggaatgg gatattcaca agaagaaatt caagaatctc ttagtaagat gaaatacgat 1440
 gaaatcacag ctacatattt gttattgggg agaaaatctt cagagctgga tgctagtgtat 1500
 tccagttcta gcagcaatct ttcacttgct aaggtaggc cgagcagtga tctcaacaac 1560
 agtactggcc agtctcctca ccacaaagtg cagagaagtg tttttcaag caaaaagcaa 1620
 agacgctaca gtgaccatgc tggaccagct attccttctg ttgtggcgta tccgaaaagg 1680
 agtcagacaa gcactgcaga tggtgacctc aaagaagatg gaatttcctc cggaaatca 1740
 agtggcagtg ctgttggagg aaagggatt gctccagcca gtcccattgt tggaatgca 1800
 agtaatccta ataaggcggg tattcctgaa cgcaagaaaa gctccactgt ccctagtagt 1860
 aacacagcat ctggtggaat gacacgacga aatacttatg tttgcagtga gagaactaca 1920
 gctgatagac actcagtgtat tcagaatggc aaagaaaaaca gcactattcc tgatcagaga 1980
 actccagttg cttcaacaca cagtatcagt agtgcagcca ccccagatcg aatccgttc 2040
 ccaagaggca ctgccagtcg tagcaatttc cacggccagc cccggaaacg gcaaccgca 2100
 acatataatg gccctcctgc ctctcccagc ctgtcccatg aagccacacc attgtccccag 2160
 actcgaagcc gaggctccac taatctttt agtaaattaa cttcaaaaact cacaaggagt 2220
 cgcaatgtat ctgctgagca aaaagatgaa aacaaagaag caaagcctcg atccctacgc 2280
 ttcacctgga gcatgaaaac cactagttca atggatcccgg gggacatgtat gcccggaaatc 2340
 cgcaaaagtgt tggacgccaa taactgcgac tatgagcaga gggagcgctt cttgtcttc 2400
 tgcgtccacg gagatgggca cgccggagaac ctgcgtcagt gggaaatgga agtgtgcaag 2460
 ctgccaagac tgtctctgaa cgggtccgg tttaagcgga tatcggggac atccatagcc 2520
 ttcaaaaata ttgcttccaa aattgccaat gagctaaagc tgtaacccag tgattatgtat 2580
 gtaaattaag tagcaagtaa agtgtttcc tgaacactga tggaaatgta tagaataata 2640
 tttaggcaat aacgtctgca tcttctaaat catgaaattt aagtctgagg acgagagc 2698

<210> 14
 <211> 2914
 <212> DNA
 <213> Homo sapiens

<400> 14
 gacggcccccgg gccaggccccgg ggatctagaa cggccgtagg gggaaaggag ccgcctccc 60

cacggcgccct	tttcggaact	gccgtggact	cgaggacgct	ggtcgcggc	ctcctagggc	120
tgtgctgttt	tgttttgcacc	ctcgattgt	gcagaattaa	agtgcagtaa	aatgtccact	180
aggaccccat	tgccaacggt	aatgaacga	gacactgaaa	accacacgtc	acatggagat	240
gggcgtcaag	aagttacctc	tcgtaccaggc	cgctcaggag	ctcggtgttag	aaactctata	300
gcctcctgtg	cagatgaaca	acctcacatc	ggaaactaca	gactgttcaa	aacaatoggc	360
aaggggaaatt	ttgcaaaaagt	aaaattggca	agacatatcc	ttacaggcag	agaggttgc	420
ataaaaaataa	ttgacaaaac	tcagttgaat	ccaacaagtc	tacaaaagct	cttcagagaa	480
gtaagaataa	tgaagatttt	aaatcatccc	aatatagtga	agttattcga	agtcattgaa	540
actcaaaaaaa	cactctacct	aatcatggaa	tatgcaagtg	gaggtaaagt	atttgactat	600
ttggttgcac	atggcaggat	gaaggaaaaaa	gaagcaagat	ctaaatttag	acagattgtg	660
tctgcagttc	aatactgcca	tcagaaacgg	atcgtacatc	gagacctca	ggctgaaaat	720
ctattgttag	atgccgatat	gaacattaaa	atagcagatt	tcggttttag	caatgaatit	780
actgttggcg	gtaaaactoga	cacgttttgt	ggcagtcctc	catacgcagc	acctgagctc	840
ttccagggca	agaaaatatga	cgggcccagaa	gtggatgtgt	ggagtctggg	ggtcatttta	900
tacacactag	tcagtggctc	acttcccttt	gatgggcaaa	acctaaagga	actgagagag	960
agagtattaa	gagggaaata	cagaattccc	ttctacatgt	ctacagactg	tgaaaacctt	1020
ctcaaacgtt	tcctggtgct	aaatccaatt	aaacgcggca	ctctagagca	aatcatgaag	1080
gacaggtgga	tcaatgcagg	gcatgaagaa	gatgaactca	aaccatttgt	tgaaccagag	1140
ctagacatct	cagaccaaaaa	aagaatagat	attatggtgg	aatgggata	ttcacaagaa	1200
gaaattcaag	aatcttttag	taagatgaaa	tacgatgaaa	tcacagctac	atatttgtt	1260
ttggggagaa	aatcttcaga	ggttaggccc	agcagtgatc	tcaacaacag	tactggccag	1320
tctcctcacc	acaaagtgca	gagaagtgtt	tcttcaagcc	aaaagcaaag	acgctacagt	1380
gaccatgctg	gaccaggat	tccttctgtt	gtggcgtatc	cgaaaaggag	tcagaccagc	1440
actgcagata	gtgacctcaa	agaagatgga	atttcctccc	ggaaatcaac	tggcagtgt	1500
gttggaggaa	agggaaattgc	tccagccagt	cccatgcttg	ggaatgcaag	taatccta	1560
aaggcggata	ttcctgaacg	caagaaaagc	tccactgtcc	ctagtagtaa	cacagcatct	1620
ggtggaatga	cacgacgaaa	tacttatgtt	tgcagtgaga	gaactacaga	tgatagacac	1680
tcagtgattc	agaatggcaa	agaaaacagc	actattcctg	atcagagaac	tccagttgt	1740
tcaacacaca	gtatcagtag	tgcagccacc	ccagatcgaa	tccgcttccc	aagaggcact	1800
gccagtcgta	gcactttcca	cggccagccc	cgggaacggc	gaaccgcaac	atataatggc	1860
cctcctgcct	ctccccagcct	gtccccatgaa	gccacaccat	tgtcccagac	tcgaagccga	1920

ggctccacta ctctctttag taaaattaact tcaaaaactca caaggagtcg	caatgtatct	1980	
gctaagcaaa aagatgaaaa caaagaagca aagcctcgat ccctacgctt	cacctggagc	2040	
atgaaaacca ctagttcaat ggatcccggg gacatgatgc gggaaatccg	caaagtgttg	2100	
gacgccaata actgcgacta tgagcagagg gagcgcattct	tgctttctg	2160	
gatgggcacg cggagaacct cgtgcagtgg gaaatggaag tgtgcaagct	gccaaagactg	2220	
tctctgaacg gggtccgggt taagcggata tcggggacat ccatagccct	caaaaatatt	2280	
gcttccaaaa ttgccaatga gctaaagctg taacccagtg attatgatgt	aaattaagta	2340	
gcaagtaaag tggtttctg aacactgatg gaaatgtata gaataatatt	taggcaataa	2400	
cgtctgcata ttctaaatca tgaaattaaa gtctgaggac gagagcacgc	ctgggagcga	2460	
aagctggcct ttttctacg aatgcactac attaaagatg tgcaacctat	gcgcffffctg	2520	
ccctacttcc gttaccctga gagtcggcgt gtggcccat ctccatgtgc	ctccccgtctg	2580	
ggtgggtgtg agagtggacg gtatgtgtgt	gaagtgggtgt atatggaagc atctccctac	2640	
actggcagcc agtcattact agtacctctg cgggagatca tccggtgcta	aaacattaca	2700	
gttgccaagg aggaaaatac tgaatgactg ctaagaatta accttaagac	cagttcatag	2760	
ttaatacagg tttacagttc atgcctgtgg ttttgtttag	gttgttttgcgtt	2820	
tgcaaaaagg	ttaaaatttat agttgtgaac attgcttgcgt	tgtgttttgcgtt	2880
cacaagataa ttaaaaattc acttttctc aggt		2914	

<210> 15
 <211> 3895
 <212> DNA
 <213> Homo sapiens

<400> 15			
ctgcaggaat tccgatcctt ccgcagggttc acctacggaa accttgcgttac	gacttttact	60	
tcctctagat agtcaagttc gaccgtcttc tcagcgctcc	gccagggccg	tgggcccggacc	120
ccggcggggc cgatccgagg gcctcaactaa accatccaat	cggtagtagc	gacggggcggt	180
gtgtacaaag ggcagggact taatcaacgc aagcttatga	cccgcaactta	ctgggaattc	240
ctcgttcatg gggataatt gcaatccccg atccccatca	cgaatgggtt	tcaacgggtt	300
acccgccct gccggcgttag ggtaggcaca cgctgagcca	gtcagtgttag	cgcgctgca	360
gccccggaca tctaaggca tcacagaccc	gttattgctc	aatctcggtt	420
cacttgcctt tctaagaagt tggggacgc	cgaccgctcg	gggtcgctgt	480
catgccagag tctcggtcg	tatcgaaatt	aaccagacaa atcgctccac	540
cggccatgca ccaccacca	cgaatcgag	aaagagctat caatctgtca	600
tgtccgggcc	gggtgagggtt	tcccggttgcgtt agtcaaatta agccgcagge	660

gtggtgtccct	tccgtcaatt	cctttaagtt	ttagctttgc	aaccataactc	ccccccggAAC	720
ccaaagactt	tggtttcccG	gaagctgccc	ggcgGGGtcAT	gggaataACG	ccgcccgcATC	780
gccgggtcggc	atcgTTtatG	gtcgGAactA	cgacGGTatC	tgatCGTtT	cgaacCTCCG	840
actttcgTTc	ttgattaatG	aaaacatttC	tggcaaATGc	tttcgctctG	gtccgtcttG	900
cgccgggtCCA	agaatttCgg	aattccgcAG	cggcggccAG	caggGCggAG	gctgaggcAG	960
caagctcgct	agagagggAG	aagcagtcgg	gCGcaggcgc	ctccTccgca	gcccgttcca	1020
tggTCggcgc	ccacagccCG	cggcggcctG	tcttgcgcTC	cacttccttc	acatcctcct	1080
ccgcctcctc	gttttcagGC	gccgcggcgc	gCGctgtgtG	gaggcccgcG	agctgaaatt	1140
cgcggTgcga	cgggagggAG	tggagaAGGA	ggtgagggGGG	cccaggatcG	cggggcgcCCC	1200
tgaggcaagg	ggacgcccGGC	gggcccgaAGC	gcagccccGCC	gcccgcaggc	tcggctccgc	1260
cactgcccGCC	ctcccggttC	cctcgctcG	gccgcccaggG	caggagAGAGA	atgagccccG	1320
ggaccCGCCG	ggggacggCC	cgggCcaggc	ccgggatcta	gacggccgta	gggggaAGGG	1380
agccGCCcTC	cccacggcgc	cttttcggAA	ctgcccgtGGA	ctcgaggacG	ctggTCggcCG	1440
gcctcctagg	gctgtgtGT	tttGTTTGA	ccctcgcatt	gtgcagaATT	aaagtgcAGT	1500
aaaatgtCCA	ctaggacccc	attgccaACG	gtgaatGAAC	gagacactGA	aaaccacacG	1560
tcacatggag	atgggcgtCA	agaagttACC	tctcgtaCC	gccgctcagg	agctcggtGT	1620
agaaaACTcta	tagcctcCTG	tgcagatGAA	caacctcaca	tcggaaACTA	cagactGTTG	1680
aaaacaatCG	gcaaggGGAA	ttttGCAAaa	gtaaaattGG	caagacatat	ccttacAGGC	1740
agagaggttG	caataaaaaAT	aattgacAAA	actcagttGA	atccaacaAG	tctacaaaAG	1800
ctcttcagAG	aagtaagaAT	aatgaagATT	ttaaatcAtC	ccaatataGT	gaagttattC	1860
gaagtcattG	aaactgaaaa	aacactctAC	ctaAtcatGG	aatatGcaAG	tggaggtGAA	1920
gtatTTGact	atTTGGTTGc	acatGGcaAG	atGAAGGAAA	aagaAGcaAG	atctaaATTT	1980
agacaggGTT	gtcaagctGG	acagactATT	aaagttaACG	tctcctttGA	tttgcttAGT	2040
ctgatgttTA	catttattGT	gtctgcAGTT	caatactGCC	atcagaaaACG	gatcgtacAT	2100
cgagacCTCA	aggctgaaaa	tctattgtta	gatGCCgata	tgaacattAA	aatAGCAGAT	2160
ttcggTTTta	gcaatGAatt	tactgttGGC	ggtaaaACTG	acacgtttG	tggcagtCCT	2220
ccatacgcAG	cacctgagCT	tttccaggGC	aagaaaATATG	acgggCCAGA	agtggatGTG	2280
tggagtctGG	gggtcatttt	atacacACTA	gtcagtggCT	cacttcctt	tgtatGGGCAA	2340
aacctaaAGG	aactgagAGA	gagagtATTA	agagggAAAT	acagaattCC	cttctacatG	2400
tctacAGACT	gtgaaaACCT	tctcaaACGT	ttcctggTGC	taaatCCAAT	taaacgcggc	2460
actctAGAGC	aaatcatGAA	ggacaggGTGG	atcaatGCAG	ggcatGAAGA	agatGAACtC	2520

aaaccatgg ttgaaccaga gctagacatc tcagacccaaa aaagaataga tattatggtg	2580
ggaatggat attcacaaga agaaattcaa gaatctctta gtaagatgaa atacgatgaa	2640
atcacagcta catatttgtt attggggaga aaatcttcag agctggatgc tagtgattcc	2700
agttctagca gcaatcttc acttgctaag gttaggccga gcagtgtatct caacaacagt	2760
actggccagt ctcctcacca caaagtgcag agaagtgttt cttcaagcca aaagcaaaga	2820
cgctacagt accatgctgg accagctatt cttctgttg tggcgtatcc gaaaaggagt	2880
cagacaagca ctgcagatgg tgacctaaaa gaagatggaa tttcctcccg gaaatcaagt	2940
ggcagtgctg ttggaggaaa gggatttgct ccagccagtc ccatgcttgg gaatgcaagt	3000
aatcctaata aggccgatat tcctgaacgc aagaaaagct ccactgtccc tagtagtaac	3060
acagcatctg gtggaatgac acgacgaaat acttatgttt gcagtgagag aactacagct	3120
gatagacact cagtgattca gaatggcaaa gaaaacagca ctattcctga tcagagaact	3180
ccagttgctt caacacacag tatcagtagt gcagccaccc cagatcgaat ccgcttccca	3240
agaggcactg ccagtcgtag cactttccac ggccagcccc gggAACGGCG aaccgcaaca	3300
tataatggcc ctccctgcctc tcccagcctg tcccattgaag ccacaccatt gtcccagact	3360
cgaagccgag gctccactaa tctcttttagt aaattaactt caaaactcac aaggagtcgc	3420
aatgtatctg ctgagcaaaa agatgaaaac aaagaagcaa agcctcgatc cctacgcttc	3480
acctggagca tgaaaaccac tagtcaatg gatccccggg acatgatgctg gggaaatccgc	3540
aaagtgttgg acgcaataa ctgcgactat gagcagaggg agcgcttctt gctcttctgc	3600
gtccacggag atgggcacgc ggagaacctc gtgcagtggg aaatggaagt gtgcaagctg	3660
ccaaagactgt ctctgaacgg ggtccggttt aagcggatat cggggacatc catggcttc	3720
aaaaatattt cttccaaaat tgccaatgag ctaaagctgt aacccagtga ttatgatgta	3780
aattaagtag caagtaaagt gttttcctga acactgatgg aaatgtataag aataatattt	3840
aggcaataac gtctgcacatc tctaaatcat gaaattaaag tctgaggacg agac	3895

<210> 16
 <211> 2145
 <212> DNA
 <213> Homo sapiens

<400> 16	
atgtccacta ggacccatt gccaacggtg aatgaacgag acactgaaaa ccacacgtca	60
catggagatg ggctgtcaaga agttacctct cgtaccagcc gtcaggagc tgggtgtaga	120
aactctatag ctcctgtgc agatgaacaa ctcacatcg gaaactacag actgttgaaa	180
acaatcggca agggaaattt tgcaaaaagta aaattggcaa gacatatcct tacaggcaga	240

gaggttgcaa taaaaataat tgacaaaact cagttgaatc caacaagtct acaaaagctc	300
ttcagagaag taagaataat gaagattttt aatcatccca atatagtgaa gttattcgaa	360
gtcattgaaa ctgaaaaaac actctaccta atcatggat atgcaagtgg aggtgaagta	420
tttgactatt tggttgcaca tggcaggatg aaggaaaaag aagcaagatc taaatttaga	480
cagattgtgt ctgcagttca atactgccat cagaaacgga tcgtacatcg agacctaag	540
gctgaaaatc tattgttaga tgccgatatg aacattaaa tagcagattt cggttttagc	600
aatgaattta ctgttggcgg taaactcgac acgttttgtg gcagtcctcc atacgcagca	660
cctgagctct tccagggcaa gaaatatgac gggccagaag tggatgtgtg gagtctgggg	720
gtcattttat acacactagt cagtggctca cttccctttg atgggcaaaa cctaaaggaa	780
ctgagagaga gagtattaag agggaaatac agaattccct tctacatgtc tacagactgt	840
gaaaaccttc tcaaacgttt cctggtgcta aatccaatta aacgcggcac tctagagcaa	900
atcatgaagg acaggtggat caatgcaggg catgaagaag atgaactcaa accatttgtt	960
gaaccagagc tagacatctc agacaaaaaa agaataagata ttatggtggg aatgggatata	1020
tcacaagaag aaattcaaga atctcttagt aagatgaaat acgatgaaat cacagctaca	1080
tatttgttat tggggagaaaa atcttcagag gttaggccga gcagtgtatct caacaacagt	1140
actggccagt ctcctcacca caaagtgcag agaagtgttt cttcaagcca aaagcaaaga	1200
cgctacagt accatgctgg accagctatt cttctgttg tggcgtatcc gaaaaggagt	1260
cagaccagca ctgcagatag tgacctcaaa gaagatggaa tttcctcccg gaaatcaagt	1320
ggcagtgctg ttggagggaaa gggaaattgct ccagccagtc ccatgcttgg gaatgcaagt	1380
aatcctaata aggcggatata tcctgaacgc aagaaaaagct ccactgtccc tagtagtaac	1440
acagcatctg gtggaatgac acgacgaaat acttatgttt gcagtgagag aactacagct	1500
gatagacact cagtgattca gaatggcaaa gaaaacagca ctattcctga tcagagaact	1560
ccagttgctt caacacacag tatcagtagt gcagccaccc cagatcgaat ccgttcccc	1620
agaggcactg ccagtcgttag cactttccac ggccagcccc gggAACGGCG aaccgcaaca	1680
tataatggcc ctctgcctc tcccagcctg tcccattgaag ccacaccatt gtcccagact	1740
cgaagccgag gctccactaa tctcttttagt aaattaactt caaaactcac aaggagtcgc	1800
aatgtatctg ctgagcaaaa agatgaaaac aaagaagcaa agcctcgatc cctacgcttc	1860
acctggagca tgaaaaccac tagttcaatg gatcccgaaa acatgatgcg ggaaatccgc	1920
aaagtgttgg acgccaataa ctgcgactat gagcagaggg agcgcttctt gctttctgc	1980
gtccacggag atgggcacgc ggagaacctc gtgcagtggg aaatggaagt gtgcaagctg	2040
ccaaagactgt ctctgaacgg ggtccggttt aagcggatata cggggacatc catagccttc	2100

aaaaatattg cttccaaaat tgccaatgag ctaaagctgt aaccc

2145

<210> 17
<211> 2193
<212> DNA
<213> Homo sapiens

<400> 17		
atgtccacta ggaccccatt gccaacggtg aatgaacgag acactgaaaa	ccacacgtca	60
catggagatg ggcgtcaaga agttacctct cgtaccagcc gctcaggagc	tcggtgtaga	120
aactctatag ctcctgtgc agatgaacaa ctcacatcg gaaactacag	actgttggaaa	180
acaatcgga aggggaattt tgcaaaagta aaattggcaa gacatatcct	tacaggcaga	240
gaggttgc当地 taaaaataat tgacaaaact cagttgaatc caacaagtct	acaaaagctc	300
ttcagagaag taagaataat gaagatttta aatcatccca atatagtgaa	gttattcgaa	360
gtcattgaaa ctgaaaaaac actctaccta atcatggaat atgcaagtgg	aggtgaagta	420
tttacttatt tggttgcaca tggcaggatg aaggaaaaag aagcaagatc	taaattttaga	480
cagattgtgt ctgcagttca atactgccat cagaaacgga tcgtacatcg	agacctcaag	540
gctgaaaatc tattgtttaga tgccgatatg aacattaaaa tagcagattt	cggtttttagc	600
aatgaatttta ctgttggcgg taaaactcgac acgttttgcgc	gcagtcctcc atacgcagca	660
cctgagctct tccagggcaa gaaatatgac gggccagaag tggatgtgtg	gagtcgtgggg	720
gtcattttat acacactagt cagtggctca cttcccttg atggcaaaaa	cctaaaggaa	780
ctgagagaga gagtattaag agggaaatac agaattccct tctacatgtc	tacagactgt	840
gaaaaccttc tcaaactgtt cctggtgcta aatccaatta aacgcggcac	tctagagcaa	900
atcatgaagg acaggtggat caatgcaggg catgaagaag atgaactcaa	accatttgg	960
gaaccagagc tagacatctc agacaaaaaa agaatacgata ttatgggg	aatggatata	1020
tcacaagaag aaattcaaga atctcttagt aagatgaaat acgtgaaat	cacagctaca	1080
tatgtttagt tggggagaaa atcttcagag ctggatgcta gtgattccag	ttctagcagc	1140
aatcttcac ttgcttaagggt taggcccggc agtgatctca acaacagtac	tggccagtct	1200
cctcaccaca aagtgcagag aagtgtttct tcaagccaaa agcaaagacg	ctacagtgcac	1260
catgctggac cagctattcc ttctgttgcgatccga aaaggagtca gaccagca	ctacagtgcac	1320
gcagatagtgcagag aagtgtttct tcaagccaaa agcaaagacg	ctacagtgcac	1380
ggaggaaaagg gaattgctcc agccagtcctc atgcttggaa atgcaagtaa	tcctaataag	1440
gcggatattc ctgaacgc当地 gaaaagctcc actgtccctta gtagtaacac	agcatctgg	1500
ggaatgacac gacgaaatac ttatgtttgc agtgagagaa ctacagctga	tagacactca	1560
gtgattcaga atggcaaaaga aaacagcact attcctgatc agagaactcc	agttgattca	1620

acacacagta tcagtagtgc agccacccca gatcgaaatcc gttcccaag aggcaactgcc	1680
agtctgtacca ctttccacgg ccagccccgg gaacggcgaa ccgcaacata taatggccct	1740
cctgccttc ccagcctgtc ccatgaagcc acaccattgt cccagactcg aagccgaggc	1800
tccactaatac tcttttagtaa attaacttca aaactcaca ggagtcgcaa tgtatctgct	1860
gagcaaaaag atgaaaacaa agaagcaaag cctcgatccc tacgcttcac ctggagcatg	1920
aaaaccacta gttcaatgga tcccggggac atgatgcggg aaatccgcaa agtgttggac	1980
gccaataact gcgactatga gcagagggag cgcttcttgc tcttctgcgt ccacggagat	2040
gggcacgcgg agaacctcggt cagtgggaa atggaagtgt gcaagctgcc aagactgtct	2100
ctgaacgggg tccggtttaa gcggatatcg gggacatcca tagccttcaa aaatattgct	2160
tccaaaatttgc ccaatgagct aaagctgtaa ccc	2193

<210> 18
<211> 3373
<212> DNA
<213> Homo sapiens

<400> 18	
caggccgcctc ctccgcagcc cgctccatgg tcggcgccca cagcccgccgg cggcctgtct	60
tgcgcgtccac ttccttcaca tcctcctccg ctcctcggtt ttcaggcgcc gccggcgccg	120
ctgtgtggag gcccgcgagc tgaaattcgc ggtgcgacgg gagggagtgg agaaggaggt	180
gagggggccc aggatcgccgg ggccgcctga ggcaaggggga cgccggcgccc ccgaagcgca	240
gcccgcgcgc cgcaggctcg gtcgcgcac tgcgcgcctc ccggcttcct cgcctcgcc	300
gcccaggcag ggagagaatg agccccggga cccgcggggg acggcccggg ccaggcccg	360
gatctagaac ggccgtaggg ggaaggagc cgcctcccc acggcgccct ttccggaaactg	420
ccgtggactc gaggacgctg gtcgcggcc tcctagggt gtgtgtttt gttttgaccc	480
tgcattgtc cagaattaaa gtgcagtaaa atgtccacta ggaccccatt gccaacggtg	540
aatgaacgag acactgaaaa ccacacgtca catggagatg ggcgtcaaga agttacctct	600
cgtaccagcc gctcaggagc tcggtgtaga aactctatag ctcctgtgc agatgaacaa	660
cctcacatcg gaaactacag actgttgaaa acaatcgca agggaaattt tgcaaaagta	720
aaattggcaa gacatatcct tacaggcaga gaggttgcaa taaaaataat tgacaaaact	780
cagttgaatc caacaagtct acaaaaagctc ttccagagaag taagaataat gaagatttt	840
aatcatccca atatagtgaa gttattcgaa gtcattgaaa ctgaaaaaac actctaccta	900
atcatggaat atgcaagtgg aggtttttttt tttgactatt tggttgcaca tggcaggatg	960
aaggaaaaag aagcaagatc taaaattttaga cagattgtgt ctgcagttca atactgccat	1020

cagaaaacgga tcgtacatcg agacctaag gctgaaaatc tattgttaga tgccgatatg	1080
aacattaaaa tagcagattt cggttttagc aatgaattta ctgttggcgg taaactcgac	1140
acgttttgtg gcagtcctcc atacgcagca cctgagctct tccagggcaa gaaatatgac	1200
ggccagaag tggatgttg gagtctgggg gtcattttat acacactgt cagtggctca	1260
cttcccttg atgggcaaaa cctaaaggaa ctgagagaga gagtattaag agggaaatac	1320
agaattccct tctacatgtc tacagactgt gaaaaccttc tcaaacfttt cctggtgcta	1380
aatccaatta aacgcggcac tctagagcaa atcatgaagg acaggtggat caatgcaggg	1440
catgaagaag atgaactcaa accatttgtt gaaccagagc tagacatctc agaccaaaaa	1500
agaatagata ttatggtggg aatggatat tcacaagaag aaattcaaga atctcttagt	1560
aagatgaaat acgatgaaat cacagctaca tatttgttac tggggagaaaa atcttcagag	1620
ctggatgcta gtgattccag ttctagcagc aatcttcac ttgctaaggt taggcccggc	1680
agtgatctca acaacagtac tggccagtct cctcaccaca aagtgcagag aagtgtttct	1740
tcaagccaaa agcaaagacg ctacagtgac catgtggac cagctattcc ttctgttgc	1800
gcgtatccga aaaggagtca gaccagcact gcagatagtgc acctcaaaga agatggaatt	1860
tcctcccgga aatcaagtgg cagtgtgtt ggaggaaagg gaattgctcc agccagtc	1920
atgcttggga atgcaagtaa tcctaataag gcggatattc ctgaacgcaa gaaaagctcc	1980
actgtcccta gtagtaaacac agcatctggt ggaatgacac gacgaaatac ttatgtttgc	2040
agtgagagaa ctacagctga tagacactca gtgattcaga atggcaaaga aaacagcact	2100
attcctgatc agagaactcc agttgttca acacacagta tcagtagtgc agccaccccc	2160
gatcaatcc gcttcccaag aggcaactgcc agtcgttagca cttccacgg ccagccccgg	2220
gaacggcgaa ccgcaacata taatggccct cctgccttc ccagcctgtc ccatgaagcc	2280
acaccattgt cccagactcg aagccgaggc tccactaatac tctttagtaa attaacttca	2340
aaactcacaat ggagaaacat gtcattcagg ttatcaaaa ggcttccaaac tgaatatgag	2400
aggaacggga gatatgaggg ctcaagtcgc aatgtatctg ctgagcaaaa agatgaaaac	2460
aaagaagcaa agcctcgatc cctacgcttc acctggagca tgaaaaccac tagttcaatg	2520
gatccccgggg acatgatgcg gggaaatccgc aaagtgttgg acgccaataa ctgcgactat	2580
gagcagaggg agcgcttctt gtcattctgc gtccacggag atgggcacgc ggagaacctc	2640
gtgcagtggg aaatggaagt gtgcaagctg ccaagactgt ctctgaacgg ggtccggttt	2700
aagcggatat cggggacatc catagccttc aaaaatattt cttccaaaat tgccaatgag	2760
ctaaagctgt aaccctgtga ttatgtatgta aattaatgtg caagtaaagt gttttccatg	2820
acactgtgg aaatgtatag aataatattt aggcaataac gtctgcacatc tctaaatcat	2880

gaaattaaag tctgaggacg agagcacgcc tgggagcgaa agctggcctt ttttctacga 2940
 atgcactaca ttaaagatgt gcaacctatg cgccccatgc cctacttccg ttacccttag 3000
 agtcggcgtg tggcccccattc tccatgtgcc tcccgatgtgg gtgggtgtga gagtggacgg 3060
 tatgtgtgtg aagtggtgta tatggaagca tctccctaca ctggcagcca gtcattacta 3120
 gtacctctgc gggagatcat ccggtgctaa aacattacag ttgccaagga ggaaaatact 3180
 gaatgactgc taagaattaa ccttaagacc agttcatagt taatacaggt ttacagttca 3240
 tgccctgtggt ttgtgttttgc ttgtttgtg tttttttagt gcaaaagggtt taaatttata 3300
 gttgtgaaca ttgtttgtgt gtgttttct aagtagattc acaagataat taaaaattca 3360
 ctttttctca ggt 3373

<210> 19
 <211> 3609
 <212> DNA
 <213> Homo sapiens

<220>
 <221> misc_feature
 <222> (3606)..(3606)
 <223> "n" is A, C, G, or T

<400> 19
 cgcctccatc cgccgcccgt tggcccggtt ccgcgcggcc tccgcggccc ccgcggccc 60
 gcctgcggc cgcggccatg ggcggccgggg tcccgatgc acggggccac taggaccctc 120
 ggcgtccctt cccctcccccc gccctgcccc ctctccggcc ggcgggaccc gggcggttctc 180
 ggcggccggc ttttggatgc ggcgtcccttgc gccggccgggg gggggggggg agagaggggg 240
 ccctgggacc cccggccccc ccacccggcc gccctgccc cccgggaccc ggagaagatg 300
 ttttcgggaa cgggtgtggc cccgggcaac gatcgaaact cggacacgc tggcaccttgc 360
 ggcagtggcc gtcctcgga caaaggcccg tcctggtcca gccgctcaact gggtgccctgt 420
 tgccggaaact ccatcgccctc ctgtcccgag gagcagcccc acgtgggcaa ctaccgcctg 480
 ctgaggacca ttgggaaagggg caactctgccc aaagtcaagc tggctcggtca catcctcaact 540
 ggtcgccggg ttgcccataa gattatcgac aaaacccaggc tgaatcccgag cagcctgcag 600
 aagctgttcc gagaagtccg catcatgaag ggcctaaacc accccaacat cgtgaagctc 660
 tttgaggtga ttgagactga gaagacgctg tacctgggtga tggagttacgc aagtgttgg 720
 gaagtgttttgc actacccgt gtcgcattggc cgcattgtggagg agaagggaaagc tcgagccaaag 780
 ttccgacaga ttgtttcggtc tgtgcactat tgccaccaga aaaatattgt acacaggggac 840
 ctgaaggctg agaacctctt gctggatgcc gaggccaaaca tcaagattgc tgactttggc 900
 ttcagcaacg agttcacgct gggatcgaaag ctggacacgt tctgcgggag ccccccattat 960

gccgccccgg agctgtttca gggcaagaag tacgacgggc cggagggtgga catctggagc	1020
ctgggagtca tcctgtacac cctcgtcagc ggctccctgc cttcgacgg gcacaacctc	1080
aaggagctgc gggagcgagt actcagaggg aagtaccggg tcccttcta catgtcaaca	1140
gactgtgaga gcattcctgag gagatttttg gtgctgaacc cagctaaacg ctgtactctc	1200
gagcaaatca tgaaaagacaa atggatcaac atcggctatg agggtgagga gttgaagcca	1260
tacacagagc ccgaggagga ctccggggac accaagagaa ttgaggtgat ggtgggtatg	1320
ggctcacacac gggaaagaaat caaagagtcc ttgaccagcc agaagtacaa cgaagtgacc	1380
gccacctacc tcctgctggg caggaagact gaggagggtg gggaccgggg cgccccaggg	1440
ctggccctgg cacgggtgcg ggcccccagc gacaccacca acggaacaag ttccagcaaa	1500
ggcaccagcc acagcaaagg gcagcggagt tcctttcca cttaccaccc ccagcgcagg	1560
catagcgatt tctgtggccc atcccctgca cccctgcacc ccaaacgcag cccgacgagc	1620
acgggggagg cggagctgaa ggaggagcgg ctgccaggcc ggaaggcggag ctgcagcacc	1680
gcggggagtg ggagtcgagg gctgcggggg tccagcccca tggtcagcag cgccccacaac	1740
cccaacaagg cagagatccc agagcggcgg aaggacagca cgagcaccccc caacaacctc	1800
cctcttagca tcatgaccccg cagaaacacc tacgtttgca cagaacgccc gggggctgag	1860
cgccctgcac ttttgcggaaa tggaaagaa aacagctcag gcacccccacg ggtgccccct	1920
gcctccccctt ccagtcacag cttggcaccc ccatcagggg agcggagccg cttggcacgc	1980
ggttccacca tccgcagcac cttccatggt ggccaggtcc gggaccggcg ggcagggggt	2040
gggggtggtg ggggtgtca gaatggccc cttgccttc ccacactggc ccatgaggct	2100
gcacccctgc cgcggggcg gcccccccc accaccaacc tttcaccaa gctgacccctc	2160
aaactgaccc gaagggttac ctcgatccc tctaaacggc agaactctaa ccgctgtgtt	2220
tcggcgcact ctctgccccca gggatccaag atcaggtcgc agacgaacct gagagaatcg	2280
ggggacctga ggtcacaagt tgcacatcac cttggatca aacggaaacc gccccccggc	2340
tgctccgatt cccctggagt gtgaagctga ccagctcgcg ccctcctgag gccctgatgg	2400
cagctctgcg ccaggccaca gcagccgccc gctggcgctg ccggccagcca cagccgttcc	2460
tgctggactg cttgcacggg ggtgcggcg ggcccgagcc cttgtcccac ttcgaagtgg	2520
aggctctgca gctgccccgg ccaggcttc ggggagttct cttccgcgt gtggcgggca	2580
ccgcctggc cttccgcacc ctcgtcaccc gcatactccaa cgacctcgcg ctctgagcc	2640
ccacggccc agggccctta ctcttccttcc ccttgcgc cttcaacttct acaggagggg	2700
aaggggccag ggaggggatt ctccctttat catcacctca gttccctga attatatttg	2760
ggggcaaaga ttgtcccttc tgctgttctc tggggccgct cagcacagaa gaaggatgag	2820

ggggctcagc	ggggggagct	ggcaccttcc	tggagcctcc	agccagtcct	gtactccctc	2880
gccctaccaa	gagggcacct	gaggagactt	tggggacagg	gcaggggcag	ggagggaaac	2940
tgagggaaatc	ttccattct	ccaaacagct	caaaaattagg	ccttgggcag	gggcagggag	3000
agctgctgag	cctaaagact	ggagaatctg	ggggactggg	agtggggtc	agagaggcag	3060
attccctccc	ctcccgccc	ctcacgctca	aacccccact	tcctgccccca	ggctggcgcg	3120
gggcactttg	tacaaatcct	tgtaaatacc	ccacaccctc	ccctctgcaa	aggtctcttg	3180
aggagctgcc	gctgtcacct	acggttttta	agttattaca	ccccgaccct	cctcctgtca	3240
gccccctcac	ctgcagcctg	ttgccccata	aatttaagag	agtccccccc	tccccaatgc	3300
tgaccctagg	atttcccttc	cctgccctca	cctgcaaatg	agttaaagaa	gaggcgtggg	3360
aatccaggca	gtggtttttc	cttcggagc	ctcggttttc	tcatctgcag	aatgggagcg	3420
gtgggggtgg	gaaggtaagg	atggtcgtgg	aagaaggcag	gatggaactc	ggcctcatcc	3480
ccgaggcccc	agttcctata	tcggggcccc	cattcatcca	ctcacactcc	cagccaccat	3540
gttacactgg	actctaagcc	acttcttact	ccagtagtaa	atttattgca	ataaacaatc	3600
attganccc						3609

<210> 20
 <211> 2085
 <212> DNA
 <213> Homo sapiens

<400> 20	agatgttctc	gccccgggtg	ctggcccccgg	gcaacgatcg	gaactcggac	acgcacatggca	60
	ccttgggcag	tggccgctcc	tcggacaaaag	gccccgtctg	gtccagccgc	tcactgggtg	120
	cccggtgccg	gaactccatc	gcctccgtc	ccgaggagca	gccccacgtg	ggcaactacc	180
	gcctgctgag	gaccattggg	aaggcaact	ttgccaaagt	caagctggct	cggcacatcc	240
	tcactggtcg	ggaggttgcc	atcaagatta	tcgacaaaac	ccagctgaat	cccagcagcc	300
	tgcagaagct	gttccgagaa	gtccgcatca	tgaaggccct	aaaccacccc	aacatcgtga	360
	agctctttga	ggtgatttag	actgagaaga	cgctgtacct	ggtgatggag	tacgcaagtg	420
	ctggagaagt	gtttgactac	ctcggtcg	atggccgcat	gaaggagaag	gaagctcgag	480
	ccaaaggccg	acagatttgtt	tcggctgtgc	actattgtca	ccagaaaaat	attgtacaca	540
	gggacctgaa	ggctgagaac	ctcttgctgg	atgccgaggc	caacatcaag	attgctgact	600
	ttggcttcag	caacgaggttc	acgctggat	cgaagctgga	cacgttctgc	gggagcccc	660
	cataatgccgc	cccggagctg	tttcagggca	agaagtacga	cgggcccggag	gtggacatct	720
	ggagcctggg	agtcatcctg	tacaccctcg	tcagcggctc	cctgcccccttc	gacgggcaca	780

acctaagga gctgcgggag cgagtactca gagggaaagta ccgggtccct ttctacatgt	840
caacagactg tgagagcatc ctgcggagat ttttgtgct gaaccaggct aaacgctgta	900
ctctcgagca aatcatgaaa gacaaatgga tcaacatcggtatgagggt gaggagttga	960
agccatacac agagccccgag gaggacttcg gggacaccaa gagaatttagt gttgtgtgg	1020
gtatgggcta cacacgggaa gaaatcaaag agtccttgac cagccagaag tacaacgaag	1080
tgaccgccac ctacccctg ctggcagga agactgagga ggggtggggac cggggcgccc	1140
cagggctggc cctggcacgg gtgcgggcgc ccagcgacac caccaacgga acaagttcca	1200
gcaaaggcac cagccacagc aaagggcagc ggagttccctt ttcacacccatc caccgcacgc	1260
gcaggcatag cgatttctgt ggccatccc ctgcacccctt gcaccccaaa cgcagccgaa	1320
cgagcacggg ggaggcggag ctgaaggagg agcggctgcc aggccgaaag gtcgagctgca	1380
gcaccgcggg gagtgggagt cgagggtgc cccctccag ccccatggtc agcagccccc	1440
acaaccccaa caaggcagag atcccgacgc ggcggaaagga cagcacgagc acccccaaca	1500
accccttcc tagcatgatg acccgacagaa acacctacgt ttgcacagaa cgcccgaaaa	1560
ctgagcgccc gtcactgttg ccaaattggaa aagaaaacag ctcaggcacc ccacgggtgc	1620
ccctgcctc cccctccagt cacagcctgg caccccccattc aggggagcgg agccgcctgg	1680
cacgcggttc caccatccgc agcaccccttcc atggtgccca ggtccgggac cggcgggac	1740
gggggtggggg tgggtgggggt gtgcagaatg ggcccccgtc ctctccaca ctggccatg	1800
aggctgcacc cctgcccccc gggcgcccccc gccccaccac caaccttcc accaagctga	1860
cctccaaact gacccgaagg gttaccctcg atccctctaa acggcagaac tctaaccgt	1920
gtgtttcggtt cgcctctctg ccccaaggat ccaagatcg gtcgcagacg aacctgagag	1980
aatcgggggaa cctgagggtca caagttgcca tctaccttgg gatcaaacgg aaaccggccc	2040
ccggctgttc cgattccctt ggagtgtgaa gtcgaccacgc tcgcg	2085

<210> 21
<211> 2278
<212> DNA
<213> Homo sapiens

<400> 21 agatgtttc gcggacggtg ctggcccccgg gcaacgatcg gaactcggac acgcattggca	60
ccttggcag tggccgcctcc tcggacaaag gcccgtccctg gtccagccgc tcactgggtg	120
cccggttgcgg gaactccatc gcctcctgtc ccgaggagca gccccacgtg ggcaactacc	180
gcctgctgag gaccattggg aaggcaact ttgcacaaatg caagctggct cggcacatcc	240
tcactggctcg ggagggttgcg atcaagatta tcgacaaaac ccagctgaat cccagcagcc	300
tgcagaagct gttccgagaa gtccgcattca tgaaggccct aaaccacccc aacatcgtga	360

agctctttga ggtgatttag actgagaaga cgctgtacctt ggtatggag tacgcaagt	420
ctggagaagt gtttactac ctcgtgtcgc atggccgcat gaaggagaag gaagctcgag	480
ccaagttccg acagattgtt tcggctgtgc actattgtca ccagaaaaat attgtacaca	540
gggacctgaa ggctgagaac ctcttgctgg atgccgaggc caacatcaag attgctgact	600
ttggcttcag caacgagttc acgctggat cgaagctgga cacgttctgc gggagcccc	660
catatgccgc cccggagctg tttcagggca agaagtacga cggggcggag gtggacatct	720
ggagcctggg agtcattcctg tacaccctcg tcagcggctc cctgcccattc gacgggcaca	780
acctcaagga gctgcggag cgagtactca gagggaaagta ccgggtccct ttctacatgt	840
caacagactg tgagagcata ctgcggagat ttttggtgct gaacccagct aaacgctgta	900
ctctcgagca aatcatgaaa gacaaatgga tcaacatcggttatgaggat gaggagttga	960
agccatacac agagcccgag gaggacttcg gggacaccaa gagaatttagt gtgatggtgg	1020
gtatggcta cacacggaa gaaatcaaag agtcattgac cagccagaag tacaacgaaag	1080
tgaccgccac ctacccctcg ctggcagga agactgagga ggggtgggac cggggcgcgg	1140
cagggctggc cctggcacgg gtgcggcgcc ccagcgacac caccaacgga acaagttcca	1200
gcaaaggcac cagccacagc aaagggcagc ggagttccctc ttccacctac caccgccagc	1260
gcaggcatag cgatttctgt ggcccatccc ctgcacccct gcaccccaaa cgcagccgaa	1320
cgagcacggg ggaggcggag ctgaaggagg agcggctgcc agggcggaaag gcgagctgca	1380
gcaccgcggg gagtgggagt cgaggctgc cccctccag cccatggc agcagccccc	1440
acaaccccaa caaggcagag atccacagc ggccggaaagga cagcacgagc acccccaaca	1500
acccctccatcc tagcatgatg acccgagaa acacctacgt ttgcacagaa cgcccgaaaa	1560
ctgagccccc gtcactgttg ccaaattggaa aagaaaacag ctcaggcacc ccacgggtgc	1620
ccctgcctc cccctccagt cacagcctgg caccccatc aggggagcgg agccgcctgg	1680
cacgcgggttc caccatccgc agcaccccttcc atggtggcca ggtccgggac cggcgggcag	1740
ggggtgtgggg tggtgtgggt gtgcagaatg ggcccccgtc ctctcccaca ctggcccatg	1800
aggctgcacc cctgcccccc gggcggcccc gccccaccac caacccatcc accaagctga	1860
cctccaaact gacccgaagg gtcgcagacg aacctgagag aatcgaaaaa cctgagggtca	1920
caagttgcca tctacccctgg gatcaaacgg aaaccgcggcc cggctgc cgttccctt	1980
ggagtgtgaa gtcgaccacg tcgcgcctc ctgaggccct gatggcagct ctgcgcagg	2040
ccacacgc cggccgcgtc cgtgcggcc agccacagcc gttccctgc gctgcctgc	2100
acgggggtgc gggcggcccc gagccctgt cccacttcga agtggaggc tgccagctga	2160
cccgccagg cttgcgggaa gttctttcc gccgtgtggc gggcaccgc ctggccctcc	2220

gcaccctcgat	cacccgcata	tccaaacgacc	tcgagctctg	agccaccacg	gtccccagg	2278
<210>	22					
<211>	4917					
<212>	DNA					
<213>	Homo sapiens					
<400>	22					
agaagatgtc	ttcgccggacg	gtgctggccc	cgggcaacga	tcggaactcg	gacacgcatt	60
gcaccttggg	cagtggccgc	tcctcggaca	aaggcccgtc	ctggtccagc	cgctcaactgg	120
gtgcccgttg	ccggaaactcc	atcgccttct	gtccccgagga	gcagccccac	gtggggcaact	180
accgcctgtt	gaggaccatt	gggaaggggca	actttgccaa	agtcaagctg	gctcggcaca	240
tcctcactgg	tcgggaggtt	gcatcaaga	ttatcgacaa	aacccagctg	aatccagca	300
gcctgcagaa	gctgttccga	gaagtccgca	tcatgaaggg	cctaaaccac	cccaacatcg	360
tgaagctctt	tgaggtgatt	gagactgaga	agacgcgtta	cctggtgatg	gagtacgcaa	420
gtgctggaga	agtgtttgac	tacctcgtgt	cgcatggccg	catgaaggag	aaggaagctc	480
gagccaagtt	ccgacagatt	gtttcggctg	tgcactattg	tcaccagaaa	aatattgtac	540
acagggacct	gaaggctgag	aacctttgc	tggatgcga	ggccaacatc	aagattgctg	600
actttggctt	cagcaacgag	ttcacgctgg	gatcgaagct	ggacacgttc	tgcgggagcc	660
ccccatatgc	cgccccggag	ctgtttcagg	gcaagaagta	cgacggggccg	gaggtggaca	720
tctggagcct	gggagtcata	ctgtacaccc	tgcgtacgg	ctccctgccc	ttcgacgggc	780
acaaccta	ggagctgcgg	gagcgagtac	tcaagaggaa	gtaccgggtc	cctttctaca	840
tgtcaacaga	ctgtgagagc	atccctgcga	gattttggt	gctgaaccca	gctaaacgct	900
gtactctcga	gcaaatacatg	aaagacaaat	ggatcaacat	cggttatgag	ggtgaggagt	960
tgaagccata	cacagagccc	gaggaggact	tcggggacac	caagagaatt	gaggtgatgg	1020
tgggtatggg	ctacacacgg	gaagaaatca	aagagtccctt	gaccagccag	aagtacaacg	1080
aagtgaccgc	cacctacctc	ctgctggca	ggaagactga	ggaggggtggg	gaccggggcg	1140
ccccaggcgt	ggccctggca	cgggtgcggg	cgccccagcga	caccaccaac	ggaacaagtt	1200
ccagcaaagg	caccagccac	agcaaagggc	agcggagttc	ctcttccacc	taccaccgac	1260
agcgcaggca	tagcgatttc	tgtggccat	cccctgcacc	cctgcacccc	aaacgcagcc	1320
cgacgaggac	gggggaggcg	gagctgaagg	aggagcggct	gccaggccgg	aaggcgagct	1380
gcagcaccgc	ggggagtggg	agtgcagggc	tgccccctc	cagccccatg	gtcagcagcg	1440
cccacaaccc	caacaaggca	gagatcccag	agcggcggaa	ggacagcacg	agcaccccca	1500
acaacctccc	tccttagcatg	atgaccccgca	gaaacaccta	cgtttgcaca	gaacgccccgg	1560

gggctgagcg cccgtcaactg ttgc当地atg ggaaagaaaa cagctcaggc accccacggg 1620
 tgccccctgc ctccccctcc agtcacagcc tggcacccccc atcaggggag cggagccgcc 1680
 tggcacgcgg ttccaccatc cgccagcacct tccatggtgg ccaggtccgg gaccgggggg 1740
 caggggtgg ggggtgggg ggtgtgcaga atgggcccccc tgc当地tccc acactggccc 1800
 atgaggctgc accccctgccc gccggggccgc cccgccccac caccAACtcc ttccaccaagc 1860
 tgacccctcaa actgacccga agggttaccc tc当地atccctc taaacggcag aactctaacc 1920
 gctgtgtttc gggccctctc ctgccccagg gatccaagat caggtcgcag acgaacctga 1980
 gagaatcggg ggacctgagg tcacaagttg cc当地tctacct tgggatcaaaa cggaaaccgc 2040
 ccccccggctg ctccgattcc cctggagtgt gaagctgacc agctcgcgcc ct当地tggaggc 2100
 cctgatggca gctctgcgcc aggccacagc agccggccgc tgccgctgcc gccagccaca 2160
 gccggttcctg ctggcctgcc tgc当地gggggg tgc当地gggggg cccgagcccc tgtcccactt 2220
 cgaagtggag gtctgccagc tgcccccggcc aggcttgcgg ggagtttctt tccgcccgtgt 2280
 ggcgggcacc cccctggact tccgc当地ctc cgtcaccgcg atctccaacg acctcgagct 2340
 ctgagccacc acggtcccag ggc当地tact cttccctctcc ct当地tgcctt tcacttctac 2400
 aggaggggaa gggccaggg aggggattct cc当地ttatca tcacctcagt ttccctgaat 2460
 tatatttggg gcaaagatt gtcccccctg ctgttctctg gggccgctca gcacagaaga 2520
 aggatgaggg ggctcagcgg gggagctgg cacccctctg gagccctccag ccagtcctgt 2580
 cctccctcgc cctaccaaga gggcacctga ggagactttg gggacagggc agggccaggg 2640
 agggaaactg aggaaatctt cc当地tccctcc caacagctca aaatttaggcc ttgggcaggg 2700
 gcagggagag ctgctgagcc taaagactgg agaatctggg ggactgggag tgggggtcag 2760
 agaggcagat tc当地tccctt cccgtccctt cacgctcaaa ccccccacttc ctgccccagg 2820
 ctggcgcggg gcactttgta caaatccttg taaatacccc acaccctccc ct当地tgc当地 2880
 gtctcttgag gagctgccgc tgtc当地cttac ggtttttaag ttattacacc cc当地ccctcc 2940
 tcctgtcagc cccctcacct gcagccgtt gccaataaa tttaagagag tcccccctcc 3000
 cccaaatgtt accctaggat ttcccttccc tgc当地tacc tgc当地atgag tt当地agaaga 3060
 ggc当地tggaa tccaggcagt ggtttttctt ttccggagccct cggttttctc atctgcagaa 3120
 tgggagccgt ggggggtggaa aggttaaggat ggtc当地ggaa gaaggcagga tggactcgg 3180
 cctcatcccc gaggccccag ttccctatatc gggccccccca ttccatccact cacactcccc 3240
 gccaccatgt tacactggac tctaaagccac ttccctactcc agtagtaaat ttattcaata 3300
 aacaatcatt gaccatgcc tactccatgc caggcccagt gctggacaca gagacatgaa 3360
 gctctgtctg tgggagacag ggattctgac acagacaccg gacaaaccat tgc当地gggg 3420

agcccagaag agaaaagtggg cagggtgggg tcattgggg aagatgctcta gaggaattaa	3480
tgctggaatg gggtgtgaa ggatgagtag gagtttagtta ggcattgagt ttgcctggg	3540
caaaaagccca gaagtgggag tatgtggtat atcttcagag aactggtaa ttcaagtgtg	3600
gctgctgtgt tggcatgga tggagaatca gcaagagaaa tgctgttata ggactaataa	3660
tccatctacg ctgcttaagc aaaaaggtat ttgttggttt atgttactta atagtccagg	3720
ggcacctggc ttcaaggtagg tttgatccag gcatcaggcc attgcatactta ttttttcagt	3780
gtaaaggtaa ttctagtaat ttttatcaag taagggctcc ttccctggtg gcacagatga	3840
cttcagcagt tagaagtttc tatccctcca gttttctgca gcagaaagac ctcattgtc	3900
agtttcccag caaaaagtccc agggcagact ctcattggcc caaatggcc atgtgatttt	3960
ctctaaacca atcactgtga ctctagagtg gccagactca gagctgcact tagtaggggt	4020
tcctcaaagg aaggtaaagt gtcatacgca ggagaaaaagg catgggagct ggacagatta	4080
tagtggttga agtctgtgca gtacagaagg gcccggctta ttccacacagc acctttgggg	4140
ccaaaatgaa taagctggac tttctccccca tggcactggg gaaccatgga agttcaggga	4200
acttcagggaa agaggcttgg tcaattcctg agagcatctt ctgtgtggg gacacagtgg	4260
taatcaagac agccccaaaca ctgccctcat agagctcaca gtccaatgga ggaggcagat	4320
gtgtcctcag gcagcgactg ggcaggcgtg gtatagggga gtccagaggt gatgcctgcc	4380
tcagccaggg agggcttctt ggaggagaag gagccagcta gacatggata ggagtgcgtt	4440
ttaggcacag caaatggcac atacaaggc cagggagcaa gagagaggac aggtcctcaa	4500
caaatggcat gtgactttgt aagttagaa ttgtgtgag gtatgggct aggggcgtca	4560
gtagggcctt gaaggttatg gacaggggccc tgggctttct tccaaaggca ctgggggagc	4620
catggcaagg ttgttaggtg ggttagagatg ggcgggtttg tgctatgtgc aggggtggaaag	4680
ggagggaagt tgacaggtca gaagatcagg aaagaggtcg gggctggaca gatggggaga	4740
gcccggatag attaaagaga gtcctgtgag gcaaaagtggg caggacctgg taacagggt	4800
ctggactgtg gctttggctg gtcagaagg tccccactgg cgtgtgtggt ctatgtagcc	4860
tctgggtgtg gagctgggat cttcaactgg ggacagtaca gtaaaagaaca tcacagc	4917

<210> 23
<211> 3226
<212> DNA
<213> Homo sapiens

<400> 23 gaccggaga agatgtcttc gcggacggtg ctggcccccgg gcaacgatcg gaactcggac	60
acgcatggca ccttgggcag tggccgctcc tcggacaaag gcccgtcctg gtccagccgc	120
tcactgggtg cccgttgcgg gaactccatc gcctcctgtc ccgaggagca gccccacgtg	180

ggcaactacc	gcctgctgag	gaccattggg	aaggccaact	ttgccaaagt	caagctggct	240
cggcacatcc	tcactggtcg	ggaggttgcc	atcaagatta	tcgacaaaac	ccagctgaat	300
cccagcagcc	tgcagaagct	gttccgagaa	gtccgcatca	tgaagggcct	aaaccacccc	360
aacatcgta	agctcttga	ggtgatttag	actgagaaga	cgctgtacct	ggtgatggag	420
tacgcaagtg	ctggagaagt	gtttgactac	ctcggtcgc	atggccgcat	gaaggagaag	480
gaagctcgag	ccaagttccg	acagattgtt	tcggctgtgc	actattgtca	ccagaaaaat	540
attgtacaca	gggacctgaa	ggctgagaac	ctttgctgg	atgccgaggc	caacatcaag	600
attgctgact	ttggcttcag	caacgagttc	acgctggat	cgaagctgga	cacgttctgc	660
gggagccccc	catatgccgc	ccggagctg	tttcagggca	agaagtacga	cgggccccggag	720
gtggacatct	ggagcctggg	agtcatcctg	tacaccctcg	tcagcggctc	cctggcccttc	780
gacgggcaca	acctcaagga	gctgcggag	cgagtactca	gagggaaagta	ccgggtccct	840
ttctacatgt	caacagactg	tgagagoatc	ctgcggagat	ttttggtgct	gaacccagct	900
aaacgctgta	ctctcgagca	aatcatgaaa	gacaaatgga	tcaacatcg	ctatgagggt	960
gaggagtta	agccatacac	agagcccgag	gaggacttcg	gggacaccaa	gagaatttag	1020
gtgatggtgg	gtatggcta	cacacggaa	gaaatcaaag	agtcccttgac	cagccagaag	1080
tacaacgaag	tgaccgecac	ctacctctg	ctgggcagga	agactgagga	gggtggggac	1140
egggggcccc	cagggctggc	cctggcacgg	gtgcgggcgc	ccagcgacac	caccaacgg	1200
acaagttcca	gcaaaggcac	cagccacagc	aaagggcagc	ggagttccctc	ttccacctac	1260
caccggccagc	gcaggcatag	cgatttctgt	ggcccatccc	ctgcacccct	gcaccccaaa	1320
cgcagcccg	cgagcacggg	ggaggcggag	ctgaaggagg	agcggctgcc	aggccggaaag	1380
gcgagctgca	gcaccgcggg	gagtggagt	cgagggctgc	ccccctccag	ccccatggtc	1440
agcagcgccc	acaacccaa	caaggcagag	atcccagagc	ggcggaaagga	cagcacgagc	1500
acccccaaca	acctccctcc	tagcatgatg	acccgcagaa	acacctacgt	ttgcacagaa	1560
cgcgggggg	ctgagcgccc	gtcactgtt	ccaaatggga	aagaaaacag	ctcaggcacc	1620
ccacgggtgc	cccctgcctc	cccctccagt	cacagcctgg	caccccatc	aggggagcgg	1680
agccgcctgg	cacgcggttc	caccatccgc	agcaccttcc	atggtgccca	ggtccgggac	1740
cggcgggcag	gggggtgggg	tgggtggggt	gtgcagaatg	ggcccoctgc	ctctcccaca	1800
ctggcccatg	aggctgcacc	cctgcccccc	ggcggcccc	gccccaccac	caacctcttc	1860
accaagctga	cctccaaact	gacccgaagg	gtcgacagacg	aacctgagag	aatcggggg	1920
cctgagggtca	caagttgcca	tctaccttgg	gatcaaacgg	aaaccggcccc	ccggctgctc	1980
cgattccctt	ggagtgtgaa	gctgaccagc	tcgcgcctc	ctgaggccct	gatggcagct	2040

ctgcgccagg ccacagcaga cggccgctgc cgatgccgcc agccacagcc gttcctgctg	2100
gcctgcctgc acgggggtgc gggcgccc gagccccctgt cccacttcga agtggaggtc	2160
tgcctcaggctgc cccggccagg ctgcggggga gttcttcc gccgtgtggc gggcacccgc	2220
ctggccttcc gcaccctcgt caccgcattc tccaacgacc tcgagctctg agccaccacg	2280
gtcccaggcc ttatcttct ctccattgtc gttcacttc tacaggaggg gaaggggcca	2340
gggagggat ttccttta tcatcaccc agtttccctg aattatattt gggggcaaag	2400
attgtccctt ctgcgtttct ctggggccgc tcagcacaga agaaggatga gggggctcag	2460
cggggggagc tggcacccctc ctggagccctc cagccagtcc tgtcctccct cgccctacca	2520
agagggcacc tgaggagact ttggggacag ggcaggggca gggagggaaa ctgagggaaat	2580
cttccattcc tcccaacagc taaaaattag gccttggca gggcagggga gagctgctga	2640
gcctaaagac tggagaatct ggggactgg gagtgggggt cagagaggca gattccttcc	2700
cctccgtcc ctcacgcgc aaaccccccac ttcctgcctt aggctggcgc ggggcacttt	2760
gtacaaatcc ttgtaaatac cccacacctt ccattctgca aaggctctttt gaggagctgc	2820
cgcgtcacc tacggttttt aagttattac accccgaccc tcctcctgtc agccccctca	2880
cgtgcagcct gttgccaat aaatttagga gagtcccccc ctcccaatg ctgaccctag	2940
gattttcctt ccctgcctc acctgcaaat gagttaaaga agaggcgtgg gaatccaggc	3000
agtggttttt ctttcggag ctcgggtttt ctcatctgca gaatgggagc ggtgggggtg	3060
ggaaggtaag gatggtcgtc caagaaggca ggatggaaact cggcctcatc cccgagggccc	3120
cagttccttat atcgggcccc ccattcatcc actcacactc ccagccacca tgttacactg	3180
gactttaagc catttcttac tccagtagta aatttattca ataaac	3226

<210> 24
<211> 745
<212> PRT
<213> Homo sapiens

<400> 24

Met Ile Arg Gly Arg Asn Ser Ala Thr Ser Ala Asp Glu Gln Pro His
1 5 10 15

Ile Gly Asn Tyr Arg Leu Leu Lys Thr Ile Gly Lys Gly Asn Phe Ala
20 25 30

Lys Val Lys Leu Ala Arg His Ile Leu Thr Gly Lys Glu Val Ala Val
35 40 45

Lys Ile Ile Asp Lys Thr Gln Leu Asn Ser Ser Leu Gln Lys Leu

50	55	60
Phe Arg Glu Val Arg Ile Met Lys Val Leu Asn His Pro Asn Ile Val		
65	70	75
Lys Leu Phe Glu Val Ile Glu Thr Glu Lys Thr Leu Tyr Leu Val Met		
85	90	95
Glu Tyr Ala Ser Gly Gly Glu Val Phe Asp Tyr Leu Val Ala His Gly		
100	105	110
Arg Met Lys Glu Lys Glu Ala Arg Ala Lys Phe Arg Gln Ile Val Ser		
115	120	125
Ala Val Gln Tyr Cys His Gln Lys Phe Ile Val His Arg Asp Leu Lys		
130	135	140
Ala Glu Asn Leu Leu Leu Asp Ala Asp Met Asn Ile Lys Ile Ala Asp		
145	150	155
Phe Gly Phe Ser Asn Glu Phe Thr Phe Gly Asn Lys Leu Asp Thr Phe		
165	170	175
Cys Gly Ser Pro Pro Tyr Ala Ala Pro Glu Leu Phe Gln Gly Lys Lys		
180	185	190
Tyr Asp Gly Pro Glu Val Asp Val Trp Ser Leu Gly Val Ile Leu Tyr		
195	200	205
Thr Leu Val Ser Gly Ser Leu Pro Phe Asp Gly Gln Asn Leu Lys Glu		
210	215	220
Leu Arg Glu Arg Val Leu Arg Gly Lys Tyr Arg Ile Pro Phe Tyr Met		
225	230	235
Ser Thr Asp Cys Glu Asn Leu Leu Lys Lys Phe Leu Ile Leu Asn Pro		
245	250	255
Ser Lys Arg Gly Thr Leu Glu Gln Ile Met Lys Asp Arg Trp Met Asn		
260	265	270
Val Gly His Glu Asp Asp Glu Leu Lys Pro Tyr Val Glu Pro Leu Pro		
275	280	285
Asp Tyr Lys Asp Pro Arg Arg Thr Glu Leu Met Val Ser Met Gly Tyr		
290	295	300

Thr Arg Glu Glu Ile Gln Asp Ser Leu Val Gly Gln Arg Tyr Asn Glu
305 310 315 320

Val Met Ala Thr Tyr Leu Leu Leu Gly Tyr Lys Ser Ser Glu Leu Glu
325 330 335

Gly Asp Thr Ile Thr Leu Lys Pro Arg Pro Ser Ala Asp Leu Thr Asn
340 345 350

Ser Ser Ala Gln Phe Pro Ser His Lys Val Gln Arg Ser Val Ser Ala
355 360 365

Asn Pro Lys Gln Arg Arg Phe Ser Asp Gln Ala Gly Pro Ala Ile Pro
370 375 380

Thr Ser Asn Ser Tyr Ser Lys Lys Thr Gln Ser Asn Asn Ala Glu Asn
385 390 395 400

Lys Arg Pro Glu Glu Asp Arg Glu Ser Gly Arg Lys Ala Ser Ser Thr
405 410 415

Ala Lys Val Pro Ala Ser Pro Leu Pro Gly Leu Glu Arg Lys Lys Thr
420 425 430

Thr Pro Thr Pro Ser Thr Asn Ser Val Leu Ser Thr Ser Thr Asn Arg
435 440 445

Ser Arg Asn Ser Pro Leu Leu Glu Arg Ala Ser Leu Gly Gln Ala Ser
450 455 460

Ile Gln Asn Gly Lys Asp Ser Leu Thr Met Pro Gly Ser Arg Ala Ser
465 470 475 480

Thr Ala Ser Ala Ser Ala Ala Val Ser Ala Ala Arg Pro Arg Gln His
485 490 495

Gln Lys Ser Met Ser Ala Ser Val His Pro Asn Lys Ala Ser Gly Leu
500 505 510

Pro Pro Thr Glu Ser Asn Cys Glu Val Pro Arg Pro Ser Thr Ala Pro
515 520 525

Gln Arg Val Pro Val Ala Ser Pro Ser Ala His Asn Ile Ser Ser Ser
530 535 540

Gly Gly Ala Pro Asp Arg Thr Asn Phe Pro Arg Gly Val Ser Ser Arg

545

550

555

560

Ser Thr Phe His Ala Gly Gln Leu Arg Gln Val Arg Asp Gln Gln Asn
565 570 575

Leu Pro Tyr Gly Val Thr Pro Ala Ser Pro Ser Gly His Ser Gln Gly
580 585 590

Arg Arg Gly Ala Ser Gly Ser Ile Phe Ser Lys Phe Thr Ser Lys Phe
595 600 605

Val Arg Arg Asn Leu Asn Glu Pro Glu Ser Lys Asp Arg Val Glu Thr
610 615 620

Leu Arg Pro His Val Val Gly Ser Gly Gly Asn Asp Lys Glu Lys Glu
625 630 635 640

Glu Phe Arg Glu Ala Lys Pro Arg Ser Leu Arg Phe Thr Trp Ser Met
645 650 655

Lys Thr Thr Ser Ser Met Glu Pro Asn Glu Met Met Arg Glu Ile Arg
660 665 670

Lys Val Leu Asp Ala Asn Ser Cys Gln Ser Glu Leu His Glu Lys Tyr
675 680 685

Met Leu Leu Cys Met His Gly Thr Pro Gly His Glu Asp Phe Val Gln
690 695 700

Trp Glu Met Glu Val Cys Lys Leu Pro Arg Leu Ser Leu Asn Gly Val
705 710 715 720

Arg Phe Lys Arg Ile Ser Gly Thr Ser Met Ala Phe Lys Asn Ile Ala
725 730 735

Ser Lys Ile Ala Asn Glu Leu Lys Leu
740 745

<210> 25

<211> 795

<212> PRT

<213> Homo sapiens

<400> 25

Met Ser Ala Arg Thr Pro Leu Pro Thr Val Asn Glu Arg Asp Thr Val
1 5 10 15

Asn His Thr Thr Val Asp Gly Tyr Thr Glu Pro His Ile Gln Pro Thr
20 25 30

Lys Ser Ser Ser Arg Gln Asn Ile Pro Arg Cys Arg Asn Ser Ile Thr
35 40 45

Ser Ala Thr Asp Glu Gln Pro His Ile Gly Asn Tyr Arg Leu Gln Lys
50 55 60

Thr Ile Gly Lys Gly Asn Phe Ala Lys Val Lys Leu Ala Arg His Val
65 70 75 80

Leu Thr Gly Arg Glu Val Ala Val Lys Ile Ile Asp Lys Thr Gln Leu
85 90 95

Asn Pro Thr Ser Leu Gln Lys Leu Phe Arg Glu Val Arg Ile Met Lys
100 105 110

Ile Leu Asn His Pro Asn Ile Val Lys Leu Phe Glu Val Ile Glu Thr
115 120 125

Glu Lys Thr Leu Tyr Leu Val Met Glu Tyr Ala Ser Gly Gly Glu Val
130 135 140

Phe Asp Tyr Leu Val Ala His Gly Arg Met Lys Glu Lys Glu Ala Arg
145 150 155 160

Ala Lys Phe Arg Gln Ile Val Ser Ala Val Gln Tyr Cys His Gln Lys
165 170 175

Tyr Ile Val His Arg Asp Leu Lys Ala Glu Asn Leu Leu Asp Gly
180 185 190

Asp Met Asn Ile Lys Ile Ala Asp Phe Gly Phe Ser Asn Glu Phe Thr
195 200 205

Val Gly Asn Lys Leu Asp Thr Phe Cys Gly Ser Pro Pro Tyr Ala Ala
210 215 220

Pro Glu Leu Phe Gln Gly Lys Tyr Asp Gly Pro Glu Val Asp Val
225 230 235 240

Trp Ser Leu Gly Val Ile Leu Tyr Thr Leu Val Ser Gly Ser Leu Pro
245 250 255

Phe Asp Gly Gln Asn Leu Lys Glu Leu Arg Glu Arg Val Leu Arg Gly
260 265 270

Lys Tyr Arg Ile Pro Phe Tyr Met Ser Thr Asp Cys Glu Asn Leu Leu
275 280 285

Lys Lys Leu Leu Val Leu Asn Pro Ile Lys Arg Gly Ser Leu Glu Gln
290 295 300

Ile Met Lys Asp Arg Trp Met Asn Val Gly His Glu Glu Glu Glu Leu
305 310 315 320

Lys Pro Tyr Thr Glu Pro Asp Pro Asp Phe Asn Asp Thr Lys Arg Ile
325 330 335

Asp Ile Met Val Thr Met Gly Phe Ala Arg Asp Glu Ile Asn Asp Ala
340 345 350

Leu Ile Asn Gln Lys Tyr Asp Glu Val Met Ala Thr Tyr Ile Leu Leu
355 360 365

Gly Arg Lys Pro Pro Glu Phe Glu Gly Gly Glu Ser Leu Ser Ser Gly
370 375 380

Asn Leu Cys Gln Arg Ser Arg Pro Ser Ser Asp Leu Asn Asn Ser Thr
385 390 395 400

Leu Gln Ser Pro Ala His Leu Lys Val Gln Arg Ser Ile Ser Ala Asn
405 410 415

Gln Lys Gln Arg Arg Phe Ser Asp His Ala Gly Pro Ser Ile Pro Pro
420 425 430

Ala Val Ser Tyr Thr Lys Arg Pro Gln Ala Asn Ser Val Glu Ser Glu
435 440 445

Gln Lys Glu Glu Trp Asp Lys Asp Val Ala Arg Lys Leu Gly Ser Thr
450 455 460

Thr Val Gly Ser Lys Ser Glu Met Thr Ala Ser Pro Leu Val Gly Pro
465 470 475 480

Glu Arg Lys Lys Ser Ser Thr Ile Pro Ser Asn Asn Val Tyr Ser Gly
485 490 495

Gly Ser Met Ala Arg Arg Asn Thr Tyr Val Cys Glu Arg Thr Thr Asp
500 505 510

Arg Tyr Val Ala Leu Gln Asn Gly Lys Asp Ser Ser Leu Thr Glu Met
515 520 525

Ser Val Ser Ser Ile Ser Ser Ala Gly Ser Ser Val Ala Ser Ala Val
530 535 540

Pro Ser Ala Arg Pro Arg His Gln Lys Ser Met Ser Thr Ser Gly His
545 550 555 560

Pro Ile Lys Val Thr Leu Pro Thr Ile Lys Asp Gly Ser Glu Ala Tyr
565 570 575

Arg Pro Gly Thr Thr Gln Arg Val Pro Ala Ala Ser Pro Ser Ala His
580 585 590

Ser Ile Ser Thr Ala Thr Pro Asp Arg Thr Arg Phe Pro Arg Gly Ser
595 600 605

Ser Ser Arg Ser Thr Phe His Gly Glu Gln Leu Arg Glu Arg Arg Ser
610 615 620

Val Ala Tyr Asn Gly Pro Pro Ala Ser Pro Ser His Glu Thr Gly Ala
625 630 635 640

Phe Ala His Ala Arg Arg Gly Thr Ser Thr Gly Ile Ile Ser Lys Ile
645 650 655

Thr Ser Lys Phe Val Arg Arg Asp Pro Ser Glu Gly Glu Ala Ser Gly
660 665 670

Arg Thr Asp Thr Ser Arg Ser Thr Ser Gly Glu Pro Lys Glu Arg Asp
675 680 685

Lys Glu Glu Gly Lys Asp Ser Lys Pro Arg Ser Leu Arg Phe Thr Trp
690 695 700

Ser Met Lys Thr Thr Ser Ser Met Asp Pro Asn Asp Met Met Arg Glu
705 710 715 720

Ile Arg Lys Val Leu Asp Ala Asn Asn Cys Asp Tyr Glu Gln Lys Glu
725 730 735

Arg Phe Leu Leu Phe Cys Val His Gly Asp Ala Arg Gln Asp Ser Leu
740 745 750

Val Gln Trp Glu Met Glu Val Cys Lys Leu Pro Arg Leu Ser Leu Asn
755 760 765

Gly Val Arg Phe Lys Arg Ile Ser Gly Thr Ser Ile Ala Phe Lys Asn
770 775 780

Ile Ala Ser Lys Ile Ala Asn Glu Leu Lys Leu
785 790 795

<210> 26
<211> 729
<212> PRT
<213> Homo sapiens

<400> 26

Met Ser Thr Arg Thr Pro Leu Pro Thr Val Asn Glu Arg Asp Thr Glu
1 5 10 15

Asn His Thr Ser His Gly Asp Gly Arg Gln Glu Val Thr Ser Arg Thr
20 25 30

Ser Arg Ser Gly Ala Arg Cys Arg Asn Ser Ile Ala Ser Cys Ala Asp
35 40 45

Glu Gln Pro His Ile Gly Asn Tyr Arg Leu Leu Lys Thr Ile Gly Lys
50 55 60

Gly Asn Phe Ala Lys Val Lys Leu Ala Arg His Ile Leu Thr Gly Arg
65 70 75 80

Glu Val Ala Ile Lys Ile Ile Asp Lys Thr Gln Leu Asn Pro Thr Ser
85 90 95

Leu Gln Lys Leu Phe Arg Glu Val Arg Ile Met Lys Ile Leu Asn His
100 105 110

Pro Asn Ile Val Lys Leu Phe Glu Val Ile Glu Thr Glu Lys Thr Leu
115 120 125

Tyr Leu Ile Met Glu Tyr Ala Ser Gly Gly Glu Val Phe Asp Tyr Leu
130 135 140

Val Ala His Gly Arg Met Lys Glu Lys Glu Ala Arg Ser Lys Phe Arg
145 150 155 160

Gln Ile Val Ser Ala Val Gln Tyr Cys His Gln Lys Arg Ile Val His
165 170 175

Arg Asp Leu Lys Ala Glu Asn Leu Leu Asp Ala Asp Met Asn Ile

180

185

190

Lys Ile Ala Asp Phe Gly Phe Ser Asn Glu Phe Thr Val Gly Gly Lys
195 200 205

Leu Asp Thr Phe Cys Gly Ser Pro Pro Tyr Ala Ala Pro Glu Leu Phe
210 215 220

Gln Gly Lys Lys Tyr Asp Gly Pro Glu Val Asp Val Trp Ser Leu Gly
225 230 235 240

Val Ile Leu Tyr Thr Leu Val Ser Gly Ser Leu Pro Phe Asp Gly Gln
245 250 255

Asn Leu Lys Glu Leu Arg Glu Arg Val Leu Arg Gly Lys Tyr Arg Ile
260 265 270

Pro Phe Tyr Met Ser Thr Asp Cys Glu Asn Leu Leu Lys Arg Phe Leu
275 280 285

Val Leu Asn Pro Ile Lys Arg Gly Thr Leu Glu Gln Ile Met Lys Asp
290 295 300

Arg Trp Ile Asn Ala Gly His Glu Glu Asp Glu Leu Lys Pro Phe Val
305 310 315 320

Glu Pro Glu Leu Asp Ile Ser Asp Gln Lys Arg Ile Asp Ile Met Val
325 330 335

Gly Met Gly Tyr Ser Gln Glu Glu Ile Gln Glu Ser Leu Ser Lys Met
340 345 350

Lys Tyr Asp Glu Ile Thr Ala Thr Tyr Leu Leu Leu Gly Arg Lys Ser
355 360 365

Ser Glu Leu Asp Ala Ser Asp Ser Ser Ser Ser Asn Leu Ser Leu
370 375 380

Ala Lys Val Arg Pro Ser Ser Asp Leu Asn Asn Ser Thr Gly Gln Ser
385 390 395 400

Pro His His Lys Val Gln Arg Ser Val Ser Ser Ser Gln Lys Gln Arg
405 410 415

Arg Tyr Ser Asp His Ala Gly Pro Ala Ile Pro Ser Val Val Ala Tyr
420 425 430

Pro Lys Arg Ser Gln Thr Ser Thr Ala Asp Gly Asp Leu Lys Glu Asp
435 440 445

Gly Ile Ser Ser Arg Lys Ser Ser Gly Ser Ala Val Gly Gly Lys Gly
450 455 460

Ile Ala Pro Ala Ser Pro Met Leu Gly Asn Ala Ser Asn Pro Asn Lys
465 470 475 480

Ala Asp Ile Pro Glu Arg Lys Lys Ser Ser Thr Val Pro Ser Ser Asn
485 490 495

Thr Ala Ser Gly Gly Met Thr Arg Arg Asn Thr Tyr Val Cys Ser Glu
500 505 510

Arg Thr Thr Ala Asp Arg His Ser Val Ile Gln Asn Gly Lys Glu Asn
515 520 525

Ser Thr Ile Pro Asp Gln Arg Thr Pro Val Ala Ser Thr His Ser Ile
530 535 540

Ser Ser Ala Ala Thr Pro Asp Arg Ile Arg Phe Pro Arg Gly Thr Ala
545 550 555 560

Ser Arg Ser Thr Phe His Gly Gln Pro Arg Glu Arg Arg Thr Ala Thr
565 570 575

Tyr Asn Gly Pro Pro Ala Ser Pro Ser Leu Ser His Glu Ala Thr Pro
580 585 590

Leu Ser Gln Thr Arg Ser Arg Gly Ser Thr Asn Leu Phe Ser Lys Leu
595 600 605

Thr Ser Lys Leu Thr Arg Ser Arg Asn Val Ser Ala Glu Gln Lys Asp
610 615 620

Glu Asn Lys Glu Ala Lys Pro Arg Ser Leu Arg Phe Thr Trp Ser Met
625 630 635 640

Lys Thr Thr Ser Ser Met Asp Pro Gly Asp Met Met Arg Glu Ile Arg
645 650 655

Lys Val Leu Asp Ala Asn Asn Cys Asp Tyr Glu Gln Arg Glu Arg Phe
660 665 670

Leu Leu Phe Cys Val His Gly Asp Gly His Ala Glu Asn Leu Val Gln

675

680

685

Trp Glu Met Glu Val Cys Lys Leu Pro Arg Leu Ser Leu Asn Gly Val
690 695 700

Arg Phe Lys Arg Ile Ser Gly Thr Ser Ile Ala Phe Lys Asn Ile Ala
705 710 715 720

Ser Lys Ile Ala Asn Glu Leu Lys Leu
725

<210> 27
<211> 713
<212> PRT
<213> Homo sapiens

<400> 27

Met Ser Thr Arg Thr Pro Leu Pro Thr Val Asn Glu Arg Asp Thr Glu
1 5 10 15

Asn His Thr Ser His Gly Asp Gly Arg Gln Glu Val Thr Ser Arg Thr
20 25 30

Ser Arg Ser Gly Ala Arg Cys Arg Asn Ser Ile Ala Ser Cys Ala Asp
35 40 45

Glu Gln Pro His Ile Gly Asn Tyr Arg Leu Leu Lys Thr Ile Gly Lys
50 55 60

Gly Asn Phe Ala Lys Val Lys Leu Ala Arg His Ile Leu Thr Gly Arg
65 70 75 80

Glu Val Ala Ile Lys Ile Ile Asp Lys Thr Gln Leu Asn Pro Thr Ser
85 90 95

Leu Gln Lys Leu Phe Arg Glu Val Arg Ile Met Lys Ile Leu Asn His
100 105 110

Pro Asn Ile Val Lys Leu Phe Glu Val Ile Glu Thr Gln Lys Thr Leu
115 120 125

Tyr Leu Ile Met Glu Tyr Ala Ser Gly Gly Lys Val Phe Asp Tyr Leu
130 135 140

Val Ala His Gly Arg Met Lys Glu Lys Glu Ala Arg Ser Lys Phe Arg
145 150 155 160

Gln Ile Val Ser Ala Val Gln Tyr Cys His Gln Lys Arg Ile Val His
165 170 175

Arg Asp Leu Lys Ala Glu Asn Leu Leu Asp Ala Asp Met Asn Ile
180 185 190

Lys Ile Ala Asp Phe Gly Phe Ser Asn Glu Phe Thr Val Gly Gly Lys
195 200 205

Leu Asp Thr Phe Cys Gly Ser Pro Pro Tyr Ala Ala Pro Glu Leu Phe
210 215 220

Gln Gly Lys Lys Tyr Asp Gly Pro Glu Val Asp Val Trp Ser Leu Gly
225 230 235 240

Val Ile Leu Tyr Thr Leu Val Ser Gly Ser Leu Pro Phe Asp Gly Gln
245 250 255

Asn Leu Lys Glu Leu Arg Glu Arg Val Leu Arg Gly Lys Tyr Arg Ile
260 265 270

Pro Phe Tyr Met Ser Thr Asp Cys Glu Asn Leu Leu Lys Arg Phe Leu
275 280 285

Val Leu Asn Pro Ile Lys Arg Gly Thr Leu Glu Gln Ile Met Lys Asp
290 295 300

Arg Trp Ile Asn Ala Gly His Glu Glu Asp Glu Leu Lys Pro Phe Val
305 310 315 320

Glu Pro Glu Leu Asp Ile Ser Asp Gln Lys Arg Ile Asp Ile Met Val
325 330 335

Gly Met Gly Tyr Ser Gln Glu Glu Ile Gln Glu Ser Leu Ser Lys Met
340 345 350

Lys Tyr Asp Glu Ile Thr Ala Thr Tyr Leu Leu Leu Gly Arg Lys Ser
355 360 365

Ser Glu Val Arg Pro Ser Ser Asp Leu Asn Asn Ser Thr Gly Gln Ser
370 375 380

Pro His His Lys Val Gln Arg Ser Val Ser Ser Ser Gln Lys Gln Arg
385 390 395 400

Arg Tyr Ser Asp His Ala Gly Pro Gly Ile Pro Ser Val Val Ala Tyr
405 410 415